Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3158982 A1 2021/06/17

(21) 3 158 982

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) **Date de dépôt PCT/PCT Filing Date:** 2020/12/10
(87) **Date publication PCT/PCT Publication Date:** 2021/06/17
(85) **Entrée phase nationale/National Entry:** 2022/05/19
(86) **N° demande PCT/PCT Application No.:** US 2020/064301
(87) **N° publication PCT/PCT Publication No.:** 2021/119304
(30) **Priorité/Priority:** 2019/12/10 (US62/946,359)

(51) **Cl.Int./Int.Cl.** ***C07K 14/195*** (2006.01), ***C07K 14/315*** (2006.01), ***C07K 14/32*** (2006.01), ***C07K 14/34*** (2006.01), ***C07K 14/39*** (2006.01), ***C07K 14/395*** (2006.01), ***C07K 14/415*** (2006.01), ***C12N 1/18*** (2006.01), ***C12P 7/06*** (2006.01)

(71) **Demandeur/Applicant:**
NOVOZYMES A/S, DK

(72) **Inventeurs/Inventors:**
TASSONE, MONICA, US;
SOPHIE CALIXTE, CA;
LIAO, SHARON, US

(74) **Agent:** WILSON LUE LLP

(54) **Titre : MICRO-ORGANISME POUR UNE FERMENTATION DE PENTOSE AMELIOREE**
(54) **Title: MICROORGANISM FOR IMPROVED PENTOSE FERMENTATION**

(57) **Abrégé/Abstract:**
Described herein are recombinant host organisms expressing a sugar transporter and an active pentose fermentation pathway. Also described are processes for producing a fermentation product, such as ethanol, from starch or cellulosic-containing material with the recombinant host organisms.

Date Submitted: 2022/05/19

CA App. No.: 3158982

Abstract:

Described herein are recombinant host organisms expressing a sugar transporter and an active pentose fermentation pathway. Also described are processes for producing a fermentation product, such as ethanol, from starch or cellulosic-containing material with the recombinant host organisms.

MICROORGANISM FOR IMPROVED PENTOSE FERMENTATION

Reference to a Sequence Listing

This application contains a Sequence Listing in computer readable form, which is incorporated herein by reference.

Background

Production of ethanol from starch and cellulosic containing materials is well-known in the art.

The most commonly industrially used commercial process for starch-containing material, often referred to as a "conventional process", includes liquefying gelatinized starch at high temperature (about 85°C) using typically a bacterial alpha-amylase, followed by simultaneous saccharification and fermentation (SSF) carried out anaerobically in the presence of typically a glucoamylase and a *Saccharomyces cerevisiae* yeast.

Yeasts which are used for production of ethanol for use as fuel, such as in the corn ethanol industry, require several characteristics to ensure cost effective production of the ethanol. These characteristics include ethanol tolerance, low by-product yield, rapid fermentation, and the ability to limit the amount of residual sugars remaining in the ferment. Such characteristics have a marked effect on the viability of the industrial process.

Yeast of the genus *Saccharomyces* exhibits many of the characteristics required for production of ethanol. In particular, strains of *Saccharomyces cerevisiae* are widely used for the production of ethanol in the fuel ethanol industry. Strains of *Saccharomyces cerevisiae* that are widely used in the fuel ethanol industry have the ability to produce high yields of ethanol under fermentation conditions found in, for example, the fermentation of corn mash. An example of such a strain is the yeast used in commercially available ethanol yeast product called ETHANOL RED®.

Efforts to establish and improve pentose (e.g., xylose) utilization of the yeast *Saccharomyces cerevisiae* have been reported (Kim *et al.*, 2013, *Biotechnol Adv.* 31(6):851-61). These include heterologous expression of xylose reductase (XR) and xylitol dehydrogenase (XDH) from naturally xylose fermenting yeasts such as *Scheffersomyces* (*Pichia*) *stipitis* and various *Candida* species, as well as the overexpression of xylulokinase (XK) and the four enzymes in the non-oxidative pentose phosphate pathway (PPP), namely transketolase (TKL), transaldolase (TAL), ribose-5-phosphate ketol-isomerase (RKI) and D-ribulose-5-phosphate 3-epimerase (RPE). Modifying the co-factor preference of *S. stipitis* XR towards NADH in such

CA 03158982 2022-5-19

systems has been found to provide metabolic advantages as well as improving anaerobic growth. Pathways replacing the XR/XDH with heterologous xylose isomerase (XI) have also been reported (e.g., WO2003/062430, WO2009/017441, WO2010/059095, WO2012/113120 and WO2012/135110). Efforts to improve arabinose utilization have been described in e.g., WO2003/095627, WO2010/074577 and US 7,977,083.

Despite improvement of ethanol production processes from cellulosic material over the past decade, uptake of pentoses (e.g., xylose, arabinose) across the yeast membrane remains a challenge. WO2019/096718 relates to variant hexose transporters which can be expressed in a yeast for improved arabinose fermentation. However, there remains a need for improved pentose sugar utilization in genetically-engineered yeast for production of bioethanol in an economically and commercially relevant scale.

Summary

Described herein are, *inter alia*, methods for producing a fermentation product, such as ethanol, from starch or cellulosic-containing material, and microorganisms suitable for use in such processes. The Applicant has surprisingly found that yeast expressing certain sugar transporters in combination with an active pentose fermentation pathway show remarkably improved utilization of pentose sugars during fermentation.

A first aspect relates to a recombinant host cell comprising a heterologous polynucleotide encoding a sugar transporter, wherein the cell comprises an active pentose fermentation pathway.

In one embodiment, the sugar transporter has an amino acid sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity, to the amino acid sequence of any one of sugar transporters described herein (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; or any one of SEQ ID NOs: 97, 116 and 138). In one embodiment, the sugar transporter differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from the amino acid sequence of any one of sugar transporters described herein (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; or any one of SEQ ID NOs: 97, 116 and 138). In one embodiment, the signal peptide comprises or consists of the amino acid sequence of any one of sugar transporters described herein (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; or any one of SEQ ID NOs: 97, 116 and 138). In one embodiment, the sugar transporter is not a transporter having a mature polypeptide sequence of

CA 03158982 2022-5-19

SEQ ID NO: 390 (or a transporter having a mature polypeptide sequence with at least 80%, e.g., at least 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to the transporter of SEQ ID NO: 390).

In one embodiment, the recombinant host cell comprises an active xylose fermentation pathway. In one embodiment, the cell comprises one or more active xylose fermentation pathway genes selected from: a heterologous polynucleotide encoding a xylose isomerase (XI), and a heterologous polynucleotide encoding a xylulokinase (XK). In one embodiment, the cell comprises one or more active xylose fermentation pathway genes selected from: a heterologous polynucleotide encoding a xylose reductase (XR), a heterologous polynucleotide encoding a xylitol dehydrogenase (XDH), and a heterologous polynucleotide encoding a xylulokinase (XK).

In one embodiment, the recombinant host cell comprises an active arabinose fermentation pathway. In one embodiment, cell comprises one or more active arabinose fermentation pathway genes selected from: a heterologous polynucleotide encoding a L-arabinose isomerase (AI), a heterologous polynucleotide encoding a L-ribulokinase (RK), and a heterologous polynucleotide encoding a L-ribulose-5-P4-epimerase (R5PE). In one embodiment, the cell comprises one or more active arabinose fermentation pathway genes selected from: a heterologous polynucleotide encoding an aldose reductase (AR), a heterologous polynucleotide encoding a L-arabinitol 4-dehydrogenase (LAD), a heterologous polynucleotide encoding a L-xylulose reductase (LXR), a heterologous polynucleotide encoding a xylitol dehydrogenase (XDH) and a heterologous polynucleotide encoding a xylulokinase (XK).

In one embodiment, the recombinant host cell comprises an active xylose fermentation pathway and an active arabinose fermentation pathway.

In one embodiment, the recombinant host cell further comprises a heterologous polynucleotide encoding a glucoamylase. In one embodiment, the glucoamylase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250.

In one embodiment, the recombinant host cell further comprises a heterologous polynucleotide encoding an alpha-amylase. In one embodiment, the alpha-amylase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256.

In one embodiment, the recombinant host cell further comprises a heterologous polynucleotide encoding a phospholipase. In one embodiment, the phospholipase has a mature

CA 03158982 2022-5-19

polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241 and 242.

In one embodiment, the recombinant host cell further comprises a heterologous polynucleotide encoding a trehalase. In one embodiment, the trehalase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 175-226.

In one embodiment, the recombinant host cell further comprises a heterologous polynucleotide encoding a protease. In one embodiment, the protease has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 9-73.

In one embodiment, the recombinant host cell further comprises a heterologous polynucleotide encoding a pullulanase. In one embodiment, the pullulanase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 114-120.

In one embodiment, the recombinant host cell is capable of higher anaerobic growth rate on pentose (e.g., xylose and/or arabinose) compared to the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2). In one embodiment, the cell is capable of higher pentose (e.g., xylose and/or arabinose) consumption compared to the same cell without the heterologous polynucleotide encoding a sugar transporter at about or after 120 hours fermentation (e.g., under conditions described in Example 2). In one embodiment, the cell is capable of consuming more than 65%, e.g., at least 70%, 75%, 80%, 85%, 90%, 95% of pentose (e.g., xylose and/or arabinose) in the medium at about or after 120 hours fermentation (e.g., under conditions described in Example 2). In one embodiment, the cell is capable of higher ethanol production compared to the same cell without the heterologous polynucleotide encoding a sugar transporter under the same conditions (e.g., after 40 hours of fermentation).

In one embodiment, the recombinant host cell further comprises a heterologous polynucleotide encoding a transketolase (TKL1). In one embodiment, the cell further comprises a heterologous polynucleotide encoding a transaldolase (TAL1). In one embodiment, the cell further comprises a disruption to an endogenous gene encoding a glycerol 3-phosphate dehydrogenase (GPD). In one embodiment, the cell further comprises a disruption to an endogenous gene encoding a glycerol 3-phosphatase (GPP).

CA 03158982 2022-5-19

In one embodiment, the recombinant host cell is a yeast cell. In one embodiment, the cell is a *Saccharomyces, Rhodotorula, Schizosaccharomyces, Kluyveromyces, Pichia, Hansenula, Rhodosporidium, Candida, Yarrowia, Lipomyces, Cryptococcus,* or *Dekkera sp.* yeast cell. In one embodiment, the cell is *Saccharomyces cerevisiae.*

A second aspect relates to methods of producing a fermentation product from a starch-containing or cellulosic-containing material, the method comprising:

(a) saccharifying the starch-containing or cellulosic-containing material; and

(b) fermenting the saccharified material of step (a) with the recombinant host cell of the first aspect.

In one embodiment, the method comprises liquefying the starch-containing material at a temperature above the initial gelatinization temperature in the presence of an alpha-amylase and/or a protease prior to saccharification. In one embodiment, the fermentation product is ethanol.

A third aspect relates to methods of producing a derivative of host cell of the first aspect, comprising culturing a host cell of the first aspect with a second host cell under conditions which permit combining of DNA between the first and second host cells, and screening or selecting for a derived host cell.

A fourth aspect relates to compositions comprising the host cell of the first aspect with one or more naturally occurring and/or non-naturally occurring components, such as components selected from the group consisting of: surfactants, emulsifiers, gums, swelling agents, and antioxidants.

Brief Description of the Figures

Figure 1 shows arabinose fermentation pathways from L-arabinose to D-xylulose 5-phosphate, which is then fermented to ethanol via the pentose phosphate pathway. The bacterial pathway utilizes genes L-arabinose isomerase (AI), L-ribulokinase (RK), and L-ribulose-5-P4-epimerase (R5PE) to convert L-arabinose to D-xylulose 5-phosphate. The fungal pathway proceeds using aldose reductase (AR), L-arabinitol 4-dehydrogenase (LAD), L-xylulose reductase (LXR), xylitol dehydrogenase (XDH) and xylulokinase (XK).

Figure 2 shows xylose fermentation pathways from D-xylose to D-xylulose 5-phosphate, which is then fermented to ethanol via the pentose phosphate pathway. The oxido-reductase pathway uses an aldolase reductase (AR, such as xylose reductase (XR)) to reduce D-xylose to xylitol followed by oxidation of xylitol to D-xylulose with xylitol dehydrogenase (XDH). The

CA 03158982 2022-5-19

isomerase pathway uses xylose isomerase (XI) to convert D-xylose directly into D-xylulose. D-xylulose is then converted to D-xylulose-5-phosphate with xylulokinase (XK).

Figure 3 shows a plasmid map for pMIBa457.

Figure 4 shows a plasmid map for pMLBA814.

Figure 5 shows a plasmid map for pMLBA647.

Figure 6 shows a plasmid map for pJDIN171.

Figure 7 shows a plasmid map for HP70.

Figure 8 shows a plasmid map for TH12.

Figure 9 shows a plasmid map for TH26.

Figure 10 shows a plasmid map for HP27.

Figure 11 shows a plasmid map for pMLBA771.

Figure 12 shows final ethanol levels from the experiment described in Example 5.

Definitions

Unless defined otherwise or clearly indicated by context, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art.

Allelic variant: The term "allelic variant" means any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

Alpha-amylase: The term "alpha amylase" means an 1,4-alpha-D-glucan glucanohydrolase, EC. 3.2.1.1, which catalyze hydrolysis of starch and other linear and branched 1,4-glucosidic oligo- and polysaccharides. For purposes of the present invention, alpha amylase activity can be determined using an alpha amylase assay described in the examples section below.

Auxiliary Activity 9: The term "Auxiliary Activity 9" or "AA9" means a polypeptide classified as a lytic polysaccharide monooxygenase (Quinlan *et al.*, 2011, *Proc. Natl. Acad. Sci. USA* 208: 15079-15084; Phillips *et al.*, 2011, *ACS Chem. Biol.* 6: 1399-1406; Lin *et al.*, 2012, *Structure* 20: 1051-1061). AA9 polypeptides were formerly classified into the glycoside hydrolase Family 61 (GH61) according to Henrissat, 1991, *Biochem. J.* 280: 309-316, and Henrissat and Bairoch, 1996, *Biochem. J.* 316: 695-696.

AA9 polypeptides enhance the hydrolysis of a cellulosic-containing material by an enzyme having cellulolytic activity. Cellulolytic enhancing activity can be determined by measuring the

CA 03158982 2022-5-19

increase in reducing sugars or the increase of the total of cellobiose and glucose from the hydrolysis of a cellulosic-containing material by cellulolytic enzyme under the following conditions: 1-50 mg of total protein/g of cellulose in pretreated corn stover (PCS), wherein total protein is comprised of 50-99.5% w/w cellulolytic enzyme protein and 0.5-50% w/w protein of an AA9 polypeptide for 1-7 days at a suitable temperature, such as 40C-80°C, e.g., 50°C, 55°C, 60°C, 65°C, or 70°C, and a suitable pH, such as 4-9, e.g., 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, or 8.5, compared to a control hydrolysis with equal total protein loading without cellulolytic enhancing activity (1-50 mg of cellulolytic protein/g of cellulose in PCS).

AA9 polypeptide enhancing activity can be determined using a mixture of CELLUCLAST® 1.5L (Novozymes A/S, Bagsværd, Denmark) and beta-glucosidase as the source of the cellulolytic activity, wherein the beta-glucosidase is present at a weight of at least 2-5% protein of the cellulase protein loading. In one embodiment, the beta-glucosidase is an *Aspergillus oryzae* beta-glucosidase (e.g., recombinantly produced in *Aspergillus oryzae* according to WO 02/095014). In another embodiment, the beta-glucosidase is an *Aspergillus fumigatus* beta-glucosidase (e.g., recombinantly produced in *Aspergillus oryzae* as described in WO 02/095014).

AA9 polypeptide enhancing activity can also be determined by incubating an AA9 polypeptide with 0.5% phosphoric acid swollen cellulose (PASC), 100 mM sodium acetate pH 5, 1 mM $MnSO_4$, 0.1% gallic acid, 0.025 mg/ml of *Aspergillus fumigatus* beta-glucosidase, and 0.01% TRITON® X-100 (4-(1,1,3,3-tetramethylbutyl)phenyl-polyethylene glycol) for 24-96 hours at 40°C followed by determination of the glucose released from the PASC.

AA9 polypeptide enhancing activity can also be determined according to WO 2013/028928 for high temperature compositions.

AA9 polypeptides enhance the hydrolysis of a cellulosic-containing material catalyzed by enzyme having cellulolytic activity by reducing the amount of cellulolytic enzyme required to reach the same degree of hydrolysis preferably at least 1.01-fold, e.g., at least 1.05-fold, at least 1.10-fold, at least 1.25-fold, at least 1.5-fold, at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, or at least 20-fold.

Beta-glucosidase: The term "beta-glucosidase" means a beta-D-glucoside glucohydrolase (E.C. 3.2.1.21) that catalyzes the hydrolysis of terminal non-reducing beta-D-glucose residues with the release of beta-D-glucose. Beta-glucosidase activity can be determined using *p*-nitrophenyl-beta-D-glucopyranoside as substrate according to the procedure of Venturi *et al.*, 2002, *J. Basic Microbiol.* 42: 55-66. One unit of beta-glucosidase is defined as 1.0 µmole of *p*-nitrophenolate anion produced per minute at 25°C, pH 4.8 from 1 mM *p*-nitrophenyl-beta-D-glucopyranoside as substrate in 50 mM sodium citrate containing 0.01% TWEEN® 20.

CA 03158982 2022-5-19

Beta-xylosidase: The term "beta-xylosidase" means a beta-D-xyloside xylohydrolase (E.C. 3.2.1.37) that catalyzes the exo-hydrolysis of short beta (1→4)-xylooligosaccharides to remove successive D-xylose residues from non-reducing termini. Beta-xylosidase activity can be determined using 1 mM *p*-nitrophenyl-beta-D-xyloside as substrate in 100 mM sodium citrate containing 0.01% TWEEN® 20 at pH 5, 40°C. One unit of beta-xylosidase is defined as 1.0 µmole of *p*-nitrophenolate anion produced per minute at 40°C, pH 5 from 1 mM *p*-nitrophenyl-beta-D-xyloside in 100 mM sodium citrate containing 0.01% TWEEN® 20.

Catalase: The term "catalase" means a hydrogen-peroxide:hydrogen-peroxide oxidoreductase (EC 1.11.1.6) that catalyzes the conversion of 2 H_2O_2 to O_2 + 2 H_2O. For purposes of the present invention, catalase activity is determined according to U.S. Patent No. 5,646,025. One unit of catalase activity equals the amount of enzyme that catalyzes the oxidation of 1 µmole of hydrogen peroxide under the assay conditions.

Catalytic domain: The term "catalytic domain" means the region of an enzyme containing the catalytic machinery of the enzyme.

Cellobiohydrolase: The term "cellobiohydrolase" means a 1,4-beta-D-glucan cellobiohydrolase (E.C. 3.2.1.91 and E.C. 3.2.1.176) that catalyzes the hydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, cellooligosaccharides, or any beta-1,4-linked glucose containing polymer, releasing cellobiose from the reducing end (cellobiohydrolase I) or non-reducing end (cellobiohydrolase II) of the chain (Teeri, 1997, *Trends in Biotechnology* 15: 160-167; Teeri *et al.*, 1998, *Biochem. Soc. Trans.* 26: 173-178). Cellobiohydrolase activity can be determined according to the procedures described by Lever *et al.*, 1972, *Anal. Biochem.* 47: 273-279; van Tilbeurgh *et al.*, 1982, *FEBS Letters* 149: 152-156; van Tilbeurgh and Claeyssens, 1985, *FEBS Letters* 187: 283-288; and Tomme *et al.*, 1988, *Eur. J. Biochem.* 170: 575-581.

Cellulolytic enzyme or cellulase: The term "cellulolytic enzyme" or "cellulase" means one or more (e.g., several) enzymes that hydrolyze a cellulosic-containing material. Such enzymes include endoglucanase(s), cellobiohydrolase(s), beta-glucosidase(s), or combinations thereof. The two basic approaches for measuring cellulolytic enzyme activity include: (1) measuring the total cellulolytic enzyme activity, and (2) measuring the individual cellulolytic enzyme activities (endoglucanases, cellobiohydrolases, and beta-glucosidases) as reviewed in Zhang *et al.*, 2006, *Biotechnology Advances* 24: 452-481. Total cellulolytic enzyme activity can be measured using insoluble substrates, including Whatman №1 filter paper, microcrystalline cellulose, bacterial cellulose, algal cellulose, cotton, pretreated lignocellulose, etc. The most common total cellulolytic activity assay is the filter paper assay using Whatman №1 filter paper

CA 03158982 2022-5-19

as the substrate. The assay was established by the International Union of Pure and Applied Chemistry (IUPAC) (Ghose, 1987, *Pure Appl. Chem.* 59: 257-68).

Cellulolytic enzyme activity can be determined by measuring the increase in production/release of sugars during hydrolysis of a cellulosic-containing material by cellulolytic enzyme(s) under the following conditions: 1-50 mg of cellulolytic enzyme protein/g of cellulose in pretreated corn stover (PCS) (or other pretreated cellulosic-containing material) for 3-7 days at a suitable temperature such as 40°C-80°C, e.g., 50°C, 55°C, 60°C, 65°C, or 70°C, and a suitable pH such as 4-9, e.g., 5.0, 5.5, 6.0, 6.5, or 7.0, compared to a control hydrolysis without addition of cellulolytic enzyme protein. Typical conditions are 1 ml reactions, washed or unwashed PCS, 5% insoluble solids (dry weight), 50 mM sodium acetate pH 5, 1 mM $MnSO_4$, 50°C, 55°C, or 60°C, 72 hours, sugar analysis by AMINEX® HPX-87H column chromatography (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

Coding sequence: The term "coding sequence" or "coding region" means a polynucleotide sequence, which specifies the amino acid sequence of a polypeptide. The boundaries of the coding sequence are generally determined by an open reading frame, which usually begins with the ATG start codon or alternative start codons such as GTG and TTG and ends with a stop codon such as TAA, TAG, and TGA. The coding sequence may be a sequence of genomic DNA, cDNA, a synthetic polynucleotide, and/or a recombinant polynucleotide.

Control sequence: The term "control sequence" means a nucleic acid sequence necessary for polypeptide expression. Control sequences may be native or foreign to the polynucleotide encoding the polypeptide, and native or foreign to each other. Such control sequences include, but are not limited to, a leader sequence, polyadenylation sequence, propeptide sequence, promoter sequence, signal peptide sequence, and transcription terminator sequence. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the polynucleotide encoding a polypeptide.

Disruption: The term "disruption" means that a coding region and/or control sequence of a referenced gene is partially or entirely modified (such as by deletion, insertion, and/or substitution of one or more nucleotides) resulting in the absence (inactivation) or decrease in expression, and/or the absence or decrease of enzyme activity of the encoded polypeptide. The effects of disruption can be measured using techniques known in the art such as detecting the absence or decrease of enzyme activity using from cell-free extract measurements referenced herein; or by the absence or decrease of corresponding mRNA (e.g., at least 25% decrease, at least 50% decrease, at least 60% decrease, at least 70% decrease, at least 80% decrease, or at

CA 03158982 2022-5-19

least 90% decrease); the absence or decrease in the amount of corresponding polypeptide having enzyme activity (e.g., at least 25% decrease, at least 50% decrease, at least 60% decrease, at least 70% decrease, at least 80% decrease, or at least 90% decrease); or the absence or decrease of the specific activity of the corresponding polypeptide having enzyme activity (e.g., at least 25% decrease, at least 50% decrease, at least 60% decrease, at least 70% decrease, at least 80% decrease, or at least 90% decrease). Disruptions of a particular gene of interest can be generated by methods known in the art, e.g., by directed homologous recombination (see *Methods in Yeast Genetics* (1997 edition), Adams, Gottschling, Kaiser, and Stems, Cold Spring Harbor Press (1998)).

Endogenous gene: The term "endogenous gene" means a gene that is native to the referenced host cell. "Endogenous gene expression" means expression of an endogenous gene.

Endoglucanase: The term "endoglucanase" means a 4-(1,3;1,4)-beta-D-glucan 4-glucanohydrolase (E.C. 3.2.1.4) that catalyzes endohydrolysis of 1,4-beta-D-glycosidic linkages in cellulose, cellulose derivatives (such as carboxymethyl cellulose and hydroxyethyl cellulose), lichenin, beta-1,4 bonds in mixed beta-1,3-1,4 glucans such as cereal beta-D-glucans or xyloglucans, and other plant material containing cellulosic components. Endoglucanase activity can be determined by measuring reduction in substrate viscosity or increase in reducing ends determined by a reducing sugar assay (Zhang *et al.*, 2006, *Biotechnology Advances* 24: 452-481). Endoglucanase activity can also be determined using carboxymethyl cellulose (CMC) as substrate according to the procedure of Ghose, 1987, *Pure and Appl. Chem.* 59: 257-268, at pH 5, 40°C.

Expression: The term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion. Expression can be measured—for example, to detect increased expression—by techniques known in the art, such as measuring levels of mRNA and/or translated polypeptide.

Expression vector: The term "expression vector" means a linear or circular DNA molecule that comprises a polynucleotide encoding a polypeptide and is operably linked to control sequences that provide for its expression.

Fermentable medium: The term "fermentable medium" or "fermentation medium" refers to a medium comprising one or more (e.g., two, several) sugars, such as glucose, fructose, sucrose, cellobiose, xylose, xylulose, arabinose, mannose, galactose, and/or soluble oligosaccharides, wherein the medium is capable, in part, of being converted (fermented) by a host cell into a desired product, such as ethanol. In some instances, the fermentation medium is

CA 03158982 2022-5-19

derived from a natural source, such as sugar cane, starch, or cellulose, and may be the result of pretreating the source by enzymatic hydrolysis (saccharification). The term fermentation medium is understood herein to refer to a medium before the fermenting organism is added, such as, a medium resulting from a saccharification process, as well as a medium used in a simultaneous saccharification and fermentation process (SSF).

Glucoamylase: The term "glucoamylase" (1,4-alpha-D-glucan glucohydrolase, EC 3.2.1.3) is defined as an enzyme that catalyzes the release of D-glucose from the non-reducing ends of starch or related oligo- and polysaccharide molecules. For purposes of the present invention, glucoamylase activity may be determined according to the procedures known in the art, such as those described in the Examples of PCT/US2019/042870, filed July 22, 2019.

Hemicellulolytic enzyme or hemicellulase: The term "hemicellulolytic enzyme" or "hemicellulase" means one or more (e.g., several) enzymes that hydrolyze a hemicellulosic material. See, for example, Shallom and Shoham, 2003, *Current Opinion In Microbiology* 6(3): 219-228). Hemicellulases are key components in the degradation of plant biomass. Examples of hemicellulases include, but are not limited to, an acetylmannan esterase, an acetylxylan esterase, an arabinanase, an arabinofuranosidase, a coumaric acid esterase, a feruloyl esterase, a galactosidase, a glucuronidase, a glucuronoyl esterase, a mannanase, a mannosidase, a xylanase, and a xylosidase. The substrates for these enzymes, hemicelluloses, are a heterogeneous group of branched and linear polysaccharides that are bound via hydrogen bonds to the cellulose microfibrils in the plant cell wall, crosslinking them into a robust network. Hemicelluloses are also covalently attached to lignin, forming together with cellulose a highly complex structure. The variable structure and organization of hemicelluloses require the concerted action of many enzymes for its complete degradation. The catalytic modules of hemicellulases are either glycoside hydrolases (GHs) that hydrolyze glycosidic bonds, or carbohydrate esterases (CEs), which hydrolyze ester linkages of acetate or ferulic acid side groups. These catalytic modules, based on homology of their primary sequence, can be assigned into GH and CE families. Some families, with an overall similar fold, can be further grouped into clans, marked alphabetically (e.g., GH-A). A most informative and updated classification of these and other carbohydrate active enzymes is available in the Carbohydrate-Active Enzymes (CAZy) database. Hemicellulolytic enzyme activities can be measured according to Ghose and Bisaria, 1987, *Pure & Appl. Chem.* 59: 1739-1752, at a suitable temperature such as 40°C-80°C, e.g., 50°C, 55°C, 60°C, 65°C, or 70°C, and a suitable pH such as 4-9, e.g., 5.0, 5.5, 6.0, 6.5, or 7.0.

Heterologous polynucleotide: The term "heterologous polynucleotide" is defined herein as a polynucleotide that is not native to the host cell; a native polynucleotide in which structural

CA 03158982 2022-5-19

modifications have been made to the coding region; a native polynucleotide whose expression is quantitatively altered as a result of a manipulation of the DNA by recombinant DNA techniques, e.g., a different (foreign) promoter; or a native polynucleotide in a host cell having one or more extra copies of the polynucleotide to quantitatively alter expression. A "heterologous gene" is a gene comprising a heterologous polynucleotide.

High stringency conditions: The term "high stringency conditions" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 micrograms/ml sheared and denatured salmon sperm DNA, and 50% formamide, following standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 0.2X SSC, 0.2% SDS at 65°C.

Host cell: The term "host cell" means any cell type that is susceptible to transformation, transfection, transduction, and the like with a nucleic acid construct or expression vector comprising a polynucleotide described herein. The term "host cell" encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication. The term "recombinant cell" is defined herein as a non-naturally occurring host cell comprising one or more (e.g., two, several) heterologous polynucleotides.

Low stringency conditions: The term "low stringency conditions" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 micrograms/ml sheared and denatured salmon sperm DNA, and 25% formamide, following standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 0.2X SSC, 0.2% SDS at 50°C.

Mature polypeptide: The term "mature polypeptide" is defined herein as a polypeptide having biological activity that is in its final form following translation and any post-translational modifications, such as N-terminal processing, C-terminal truncation, glycosylation, phosphorylation, etc. The mature polypeptide sequence lacks a signal sequence, which may be determined using techniques known in the art (*See*, e.g., Zhang and Henzel, 2004, *Protein Science* 13: 2819-2824). The term "mature polypeptide coding sequence" means a polynucleotide that encodes a mature polypeptide.

Medium stringency conditions: The term "medium stringency conditions" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 micrograms/ml sheared and denatured salmon sperm DNA, and 35% formamide, following standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 0.2X SSC, 0.2% SDS at 55°C.

CA 03158982 2022-5-19

Medium-high stringency conditions: The term "medium-high stringency conditions" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 micrograms/ml sheared and denatured salmon sperm DNA, and 35% formamide, following standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 0.2X SSC, 0.2% SDS at 60°C.

Nucleic acid construct: The term "nucleic acid construct" means a polynucleotide comprises one or more (e.g., two, several) control sequences. The polynucleotide may be single-stranded or double-stranded, and may be isolated from a naturally occurring gene, modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature, or synthetic.

Operably linked: The term "operably linked" means a configuration in which a control sequence is placed at an appropriate position relative to the coding sequence of a polynucleotide such that the control sequence directs expression of the coding sequence.

Pentose: The term "pentose" means a five-carbon monosaccharide (e.g., xylose, arabinose, ribose, lyxose, ribulose, and xylulose). Pentoses, such as D-xylose and L-arabinose, may be derived, e.g., through saccharification of a plant cell wall polysaccharide.

Active pentose fermentation pathway: As used herein, a host cell or fermenting organism having an "active pentose fermentation pathway" produces active enzymes necessary to catalyze each reaction of a metabolic pathway in a sufficient amount to produce a fermentation product (e.g., ethanol) from pentose, and therefore is capable of producing the fermentation product in measurable yields when cultivated under fermentation conditions in the presence of pentose. A host cell or fermenting organism having an active pentose fermentation pathway comprises one or more active pentose fermentation pathway genes. A "pentose fermentation pathway gene" as used herein refers to a gene that encodes an enzyme involved in an active pentose fermentation pathway. In some embodiments, the active pentose fermentation pathway is an "active xylose fermentation pathway" (ie produces a fermentation product, such as ethanol, from xylose) or an "active arabinose fermentation pathway (ie produces a fermentation product, such as ethanol, from arabinose).

The active enzymes necessary to catalyze each reaction in an active pentose fermentation pathway may result from activities of endogenous gene expression, activities of heterologous gene expression, or from a combination of activities of endogenous and heterologous gene expression, as described in more detail herein.

Phospholipase: The term "phospholipase" means an enzyme that catalyzes the conversion of phospholipids into fatty acids and other lipophilic substances, such as

CA 03158982 2022-5-19

phospholipase A (EC numbers 3.1.1.4, 3.1.1.5 and 3.1.1.32) or phospholipase C (EC numbers 3.1.4.3 and 3.1.4.11). Phospholipase activity may be determined using activity assays known in the art.

Pretreated corn stover: The term "Pretreated Corn Stover" or "PCS" means a cellulosic-containing material derived from corn stover by treatment with heat and dilute sulfuric acid, alkaline pretreatment, neutral pretreatment, or any pretreatment known in the art.

Protease: The term "protease" is defined herein as an enzyme that hydrolyses peptide bonds. It includes any enzyme belonging to the EC 3.4 enzyme group (including each of the thirteen subclasses thereof). The EC number refers to Enzyme Nomenclature 1992 from NC-IUBMB, Academic Press, San Diego, California, including supplements 1-5 published in *Eur. J. Biochem.* 223: 1-5 (1994); *Eur. J. Biochem.* 232: 1-6 (1995); *Eur. J. Biochem.* 237: 1-5 (1996); *Eur. J. Biochem.* 250: 1-6 (1997); and *Eur. J. Biochem.* 264: 610-650 (1999); respectively. The term "subtilases" refer to a sub-group of serine protease according to Siezen et al., 1991, *Protein Engng.* 4: 719-737 and Siezen et al., 1997, *Protein Science* 6: 501-523. Serine proteases or serine peptidases is a subgroup of proteases characterised by having a serine in the active site, which forms a covalent adduct with the substrate. Further the subtilases (and the serine proteases) are characterised by having two active site amino acid residues apart from the serine, namely a histidine and an aspartic acid residue. The subtilases may be divided into 6 sub-divisions, i.e. the Subtilisin family, the Thermitase family, the Proteinase K family, the Lantibiotic peptidase family, the Kexin family and the Pyrolysin family. The term "protease activity" means a proteolytic activity (EC 3.4). Protease activity may be determined using methods described in the art (e.g., US 2015/0125925) or using commercially available assay kits (e.g., Sigma-Aldrich).

Pullulanase: The term "pullulanase" means a starch debranching enzyme having pullulan 6-glucano-hydrolase activity (EC 3.2.1.41) that catalyzes the hydrolysis the α-1,6-glycosidic bonds in pullulan, releasing maltotriose with reducing carbohydrate ends. For purposes of the present invention, pullulanase activity can be determined according to a PHADEBAS assay or the sweet potato starch assay described in WO2016/087237.

Sequence Identity: The relatedness between two amino acid sequences or between two nucleotide sequences is described by the parameter "sequence identity".

For purposes described herein, the degree of sequence identity between two amino acid sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, *J. Mol. Biol.* **1970**, *48*, 443-453) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al., *Trends Genet* **2000**, *16*, 276-277), preferably version 3.0.0 or later. The optional parameters used are gap open

CA 03158982 2022-5-19

penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle labeled "longest identity" (obtained using the –nobrief option) is used as the percent identity and is calculated as follows:

(Identical Residues x 100)/(Length of the Referenced Sequence – Total Number of Gaps in Alignment)

For purposes described herein, the degree of sequence identity between two deoxyribonucleotide sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, **1970**, *supra*) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al., 2000, *supra*), preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EDNAFULL (EMBOSS version of NCBI NUC4.4) substitution matrix. The output of Needle labeled "longest identity" (obtained using the –nobrief option) is used as the percent identity and is calculated as follows:

(Identical Deoxyribonucleotides x 100)/(Length of Referenced Sequence – Total Number of Gaps in Alignment)

Signal peptide: The term "signal peptide" is defined herein as a peptide linked (fused) in frame to the amino terminus of a polypeptide having biological activity and directs the polypeptide into the cell's secretory pathway. Signal sequences may be determined using techniques known in the art (*See*, e.g., Zhang and Henzel, 2004, *Protein Science* 13: 2819-2824). The polypeptides described herein may comprise any suitable signal peptide known in the art, or any signal peptide described in U.S. Provisional application No. 62/883,519, filed August 6, 2019 (incorporated herein by reference).

Trehalase: The term "trehalase" means an enzyme which degrades trehalose into its unit monosaccharides (i.e., glucose). Trehalases are classified in EC 3.2.1.28 (alpha,alpha-trehalase) and EC. 3.2.1.93 (alpha,alpha-phosphotrehalase). The EC classes are based on recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB). Description of EC classes can be found on the internet, e.g., on "http://www.expasy.org/enzyme/". Trehalases are enzymes that catalyze the following reactions:

EC 3.2.1.28: Alpha,alpha-trehalose + $H_2O \Leftrightarrow$ 2 D-glucose;

EC 3.2.1. 93: Alpha,alpha-trehalose 6-phosphate + $H_2O \Leftrightarrow$ D-glucose + D-glucose 6-phosphate.

Trehalase activity may be determined according to procedures known in the art.

Very high stringency conditions: The term "very high stringency conditions" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X

CA 03158982 2022-5-19

SSPE, 0.3% SDS, 200 micrograms/ml sheared and denatured salmon sperm DNA, and 50% formamide, following standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 0.2X SSC, 0.2% SDS at 70°C.

Very low stringency conditions: The term "very low stringency conditions" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 micrograms/ml sheared and denatured salmon sperm DNA, and 25% formamide, following standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 0.2X SSC, 0.2% SDS at 45°C.

Xylanase: The term "xylanase" means a 1,4-beta-D-xylan-xylohydrolase (E.C. 3.2.1.8) that catalyzes the endohydrolysis of 1,4-beta-D-xylosidic linkages in xylans. Xylanase activity can be determined with 0.2% AZCL-arabinoxylan as substrate in 0.01% TRITON® X-100 and 200 mM sodium phosphate pH 6 at 37°C. One unit of xylanase activity is defined as 1.0 µmole of azurine produced per minute at 37°C, pH 6 from 0.2% AZCL-arabinoxylan as substrate in 200 mM sodium phosphate pH 6.

Xylose Isomerase: The term "Xylose Isomerase" or "XI" means an enzyme which can catalyze D-xylose into D-xylulose *in vivo*, and convert D-glucose into D-fructose in vitro. Xylose isomerase is also known as "glucose isomerase" and is classified as E.C. 5.3.1.5. As the structure of the enzyme is very stable, the xylose isomerase is a good model for studying the relationships between protein structure and functions (Karimaki et al., Protein Eng Des Sel, 12004, 17 (12):861-869). Xylose Isomerase activity may be determined using techniques known in the art (e.g., a coupled enzyme assay using D-sorbitol dehygrogenase, as described by Verhoeven et. al., 2017, *Sci Rep* 7, 46155).

Reference to "about" a value or parameter herein includes embodiments that are directed to that value or parameter *per se*. For example, description referring to "about X" includes the embodiment "X". When used in combination with measured values, "about" includes a range that encompasses at least the uncertainty associated with the method of measuring the particular value, and can include a range of plus or minus two standard deviations around the stated value.

Likewise, reference to a gene or polypeptide that is "derived from" another gene or polypeptide X, includes the gene or polypeptide X.

As used herein and in the appended claims, the singular forms "a," "or," and "the" include plural referents unless the context clearly dictates otherwise.

It is understood that the embodiments described herein include "consisting" and/or "consisting essentially of" embodiments. As used herein, except where the context requires otherwise due to express language or necessary implication, the word "comprise" or variations

CA 03158982 2022-5-19

such as "comprises" or "comprising" is used in an inclusive sense, i.e. to specify the presence of the stated features but not to preclude the presence or addition of further features in various embodiments.

DETAILED DESCRIPTION

Described herein, *inter alia*, are host cells/fermention organism, and methods for producing a fermentation product, such as ethanol, from starch or cellulosic containing material. The Applicant has surprisingly found that yeast expressing certain sugar transporters in combination with an active pentose fermentation pathway show remarkably improved utilization of pentose sugars during fermentation.

In one aspect is a method of producing a fermentation product from a starch-containing or cellulosic-containing material comprising:

(a) saccharifying the starch-containing or cellulosic-containing material; and

(b) fermenting the saccharified material of step (a) with a recombinant host cell;

wherein the host cell comprises an active pentose fermentation pathway and a sugar transporter.

Steps a) and b) may be carried out either sequentially or simultaneously (SSF). In one embodiment, steps a) and b) are carried out simultaneously (SSF). In another embodiment, steps a) and b) are carried out sequentially.

In some embodiments of the methods described herein, fermentation of step (b) consumes a greater amount of pentose (e.g., xylose and/or arabinose) e.g., at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 75% or 90% more when compared to the method using the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2). In some embodiments, more than 65%, e.g., at least 70%, 75%, 80%, 85%, 90%, 95% of pentose (e.g., xylose and/or arabinose) in the medium is consumed.

Host cells and Fermenting organisms

The host cells and fermenting organisms described herein may be derived from any host cell known to the skilled artisan, such as a cell capable of producing a fermentation product (e.g., ethanol). As used herein, a "derivative" of strain is derived from a referenced strain, such as through mutagenesis, recombinant DNA technology, mating, cell fusion, or cytoduction between yeast strains. Those skilled in the art will understand that the genetic alterations, including metabolic modifications exemplified herein, may be described with reference to a suitable host

CA 03158982 2022-5-19

organism and their corresponding metabolic reactions or a suitable source organism for desired genetic material such as genes for a desired metabolic pathway. However, given the complete genome sequencing of a wide variety of organisms and the high level of skill in the area of genomics, those skilled in the art can apply the teachings and guidance provided herein to other organisms. For example, the metabolic alterations exemplified herein can readily be applied to other species by incorporating the same or analogous encoding nucleic acid from species other than the referenced species.

The host cells described herein can be from any suitable host, such as a yeast strain, including, but not limited to, a *Saccharomyces*, *Rhodotorula*, *Schizosaccharomyces*, *Kluyveromyces*, *Pichia*, *Hansenula*, *Rhodosporidium*, *Candida*, *Yarrowia*, *Lipomyces*, *Cryptococcus*, or *Dekkera sp.* cell. In particular, *Saccharomyces* host cells are contemplated, such as *Saccharomyces cerevisiae*, *bayanus* or *carlsbergensis* cells. Preferably, the yeast cell is a *Saccharomyces cerevisiae* cell. Suitable cells can, for example, be derived from commercially available strains and polyploid or aneuploid industrial strains, including but not limited to those from Superstart™, THERMOSACC®, C5 FUEL™, XyloFerm®, etc. (Lallemand); RED STAR and ETHANOL RED® (Fermentis/Lesaffre); FALI (AB Mauri); Baker's Best Yeast, Baker's Compressed Yeast, etc. (Fleishmann's Yeast); BIOFERM AFT, XP, CF, and XR (North American Bioproducts Corp.); Turbo Yeast (Gert Strand AB); and FERMIOL® (DSM Specialties). Other useful yeast strains are available from biological depositories such as the American Type Culture Collection (ATCC) or the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), such as, e.g., BY4741 (e.g., ATCC 201388); Y108-1 (ATCC PTA.10567) and NRRL YB-1952 (ARS Culture Collection). Still other *S. cerevisiae* strains suitable as host cells DBY746, [Alpha][Eta]22, S150-2B, GPY55-15Ba, CEN.PK, USM21, TMB3500, TMB3400, VTT-A-63015, VTT-A-85068, VTT-c-79093 and their derivatives as well as Saccharomyces sp. 1400, 424A (LNH-ST), 259A (LNH-ST) and derivatives thereof. In one embodiment, the recombinant cell is a derivative of a strain *Saccharomyces cerevisiae* CIBTS1260 (deposited under Accession No. NRRL Y-50973 at the Agricultural Research Service Culture Collection (NRRL), Illinois 61604 U.S.A.).

The host cell or fermenting organism may be *Saccharomyces* strain, e.g., *Saccharomyces cerevisiae* strain produced using the method described and concerned in US patent no. 8,257,959-BB.

The strain may also be a derivative of *Saccharomyces cerevisiae* strain NMI V14/004037 (*See*, WO2015/143324 and WO2015/143317 each incorporated herein by reference), strain nos. V15/004035, V15/004036, and V15/004037 (*See*, WO 2016/153924 incorporated herein by

CA 03158982 2022-5-19

reference), strain nos. V15/001459, V15/001460, V15/001461 (*See*, WO2016/138437 incorporated herein by reference), strain no. NRRL Y67342 (*See*, WO2018/098381 incorporated herein by reference), strain nos. NRRL Y67549 and NRRL Y67700 (*See*, PCT/US2019/018249 incorporated herein by reference), or any strain described in WO2017/087330 (incorporated herein by reference).

The fermenting organisms according to the invention have been generated in order to, e.g., improve fermentation yield and to improve process economy by cutting enzyme costs since part or all of the necessary enzymes needed to improve method performance are be produced by the fermenting organism.

The host cells and fermenting organisms described herein may utilize expression vectors comprising the coding sequence of one or more (e.g., two, several) heterologous genes linked to one or more control sequences that direct expression in a suitable cell under conditions compatible with the control sequence(s). Such expression vectors may be used in any of the cells and methods described herein. The polynucleotides described herein may be manipulated in a variety of ways to provide for expression of a desired polypeptide. Manipulation of the polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides utilizing recombinant DNA methods are well known in the art.

A construct or vector (or multiple constructs or vectors) comprising the one or more (e.g., two, several) heterologous genes may be introduced into a cell so that the construct or vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector as described earlier.

The various nucleotide and control sequences may be joined together to produce a recombinant expression vector that may include one or more (e.g., two, several) convenient restriction sites to allow for insertion or substitution of the polynucleotide at such sites. Alternatively, the polynucleotide(s) may be expressed by inserting the polynucleotide(s) or a nucleic acid construct comprising the sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

The recombinant expression vector may be any vector (e.g., a plasmid or virus) that can be conveniently subjected to recombinant DNA procedures and can bring about expression of the polynucleotide. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vector may be a linear or closed circular plasmid.

CA 03158982 2022-5-19

The vector may be an autonomously replicating vector, i.e., a vector that 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 that, when introduced into the host cell, 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 that together contain the total DNA to be introduced into the genome of the cell, or a transposon, may be used.

The expression vector may contain any suitable promoter sequence that is recognized by a cell for expression of a gene described herein. The promoter sequence contains transcriptional control sequences that mediate the expression of the polypeptide. The promoter may be any polynucleotide that shows transcriptional activity in the cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the cell.

Each heterologous polynucleotide described herein may be operably linked to a promoter that is foreign to the polynucleotide. For example, in one embodiment, the nucleic acid construct encoding the fusion protein is operably linked to a promoter foreign to the polynucleotide. The promoters may be identical to or share a high degree of sequence identity (e.g., at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%) with a selected native promoter.

Examples of suitable promoters for directing the transcription of the nucleic acid constructs in a yeast cells, include, but are not limited to, the promoters obtained from the genes for enolase, (e.g., *S. cerevisiae* enolase or *I. orientalis* enolase (ENO1)), galactokinase (e.g., *S. cerevisiae* galactokinase or *I. orientalis* galactokinase (GAL1)), alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (e.g., *S. cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase or *I. orientalis* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH1, ADH2/GAP)), triose phosphate isomerase (e.g., *S. cerevisiae* triose phosphate isomerase or *I. orientalis* triose phosphate isomerase (TPI)), metallothionein (e.g., *S. cerevisiae* metallothionein or *I. orientalis* metallothionein (CUP1)), 3-phosphoglycerate kinase (e.g., *S. cerevisiae* 3-phosphoglycerate kinase or *I. orientalis* 3-phosphoglycerate kinase (PGK)), PDC1, xylose reductase (XR), xylitol dehydrogenase (XDH), L-(+)-lactate-cytochrome c oxidoreductase (CYB2), translation elongation factor-1 (TEF1), translation elongation factor-2 (TEF2), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), and orotidine 5'-phosphate decarboxylase (URA3) genes. Other suitable promoters may be obtained from *S. cerevisiae*

CA 03158982 2022-5-19

TDH3, HXT7, PGK1, RPL18B and CCW12 genes. Additional useful promoters for yeast host cells are described by Romanos *et al.*, 1992, *Yeast* 8: 423-488.

The control sequence may also be a suitable transcription terminator sequence, which is recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3'-terminus of the polynucleotide encoding the polypeptide. Any terminator that is functional in the yeast cell of choice may be used. The terminator may be identical to or share a high degree of sequence identity (e.g., at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%) with the selected native terminator.

Suitable terminators for yeast host cells may be obtained from the genes for enolase (e.g., *S. cerevisiae* or *I. orientalis* enolase cytochrome C (e.g., *S. cerevisiae* or *I. orientalis* cytochrome (CYC1)), glyceraldehyde-3-phosphate dehydrogenase (e.g., *S. cerevisiae* or *I. orientalis* glyceraldehyde-3-phosphate dehydrogenase (gpd)), PDC1, XR, XDH, transaldolase (TAL), transketolase (TKL), ribose 5-phosphate ketol-isomerase (RKI), CYB2, and the galactose family of genes (especially the GAL10 terminator). Other suitable terminators may be obtained from *S. cerevisiae* ENO2 or TEF1 genes. Additional useful terminators for yeast host cells are described by Romanos *et al.*, 1992, *supra*.

The control sequence may also be an mRNA stabilizer region downstream of a promoter and upstream of the coding sequence of a gene which increases expression of the gene.

Examples of suitable mRNA stabilizer regions are obtained from a *Bacillus thuringiensis cryIIIA* gene (WO 94/25612) and a *Bacillus subtilis* SP82 gene (Hue et al., 1995, *Journal of Bacteriology* 177: 3465-3471).

The control sequence may also be a suitable leader sequence, when transcribed is a non-translated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5'-terminus of the polynucleotide encoding the polypeptide. Any leader sequence that is functional in the yeast cell of choice may be used.

Suitable leaders for yeast host cells are obtained from the genes for enolase (e.g., *S. cerevisiae* or *I. orientalis* enolase (ENO-1)), 3-phosphoglycerate kinase (e.g., *S. cerevisiae* or *I. orientalis* 3-phosphoglycerate kinase), alpha-factor (e.g., *S. cerevisiae* or *I. orientalis* alpha-factor), and alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (e.g., *S. cerevisiae* or *I. orientalis* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP)).

The control sequence may also be a polyadenylation sequence; a sequence operably linked to the 3'-terminus of the polynucleotide and, when transcribed, is recognized by the host

CA 03158982 2022-5-19

cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence that is functional in the host cell of choice may be used. Useful polyadenylation sequences for yeast cells are described by Guo and Sherman, 1995, *Mol. Cellular Biol.* 15: 5983-5990.

The control sequence may also be a signal peptide coding region that encodes a signal peptide linked to the N-terminus of a polypeptide and directs the polypeptide into the cell's secretory pathway. The 5'-end of the coding sequence of the polynucleotide may inherently contain a signal peptide coding sequence naturally linked in translation reading frame with the segment of the coding sequence that encodes the polypeptide. Alternatively, the 5'-end of the coding sequence may contain a signal peptide coding sequence that is foreign to the coding sequence. A foreign signal peptide coding sequence may be required where the coding sequence does not naturally contain a signal peptide coding sequence. Alternatively, a foreign signal peptide coding sequence may simply replace the natural signal peptide coding sequence in order to enhance secretion of the polypeptide. However, any signal peptide coding sequence that directs the expressed polypeptide into the secretory pathway of a host cell may be used. Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding sequences are described by Romanos *et al.*, 1992, *supra.*

The control sequence may also be a propeptide coding sequence that encodes a propeptide positioned at the N-terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to an active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding sequence may be obtained from the genes for *Bacillus subtilis* alkaline protease (*aprE*), *Bacillus subtilis* neutral protease (*nprT*), *Myceliophthora thermophila* laccase (WO 95/33836), *Rhizomucor miehei* aspartic proteinase, and *Saccharomyces cerevisiae* alpha-factor.

Where both signal peptide and propeptide sequences are present, the propeptide sequence is positioned next to the N-terminus of a polypeptide and the signal peptide sequence is positioned next to the N-terminus of the propeptide sequence.

It may also be desirable to add regulatory sequences that allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those that cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory

CA 03158982 2022-5-19

systems in prokaryotic systems include the *lac*, *tac*, and *trp* operator systems. In yeast, the ADH2 system or GAL1 system may be used.

The vectors may contain one or more (e.g., two, several) selectable markers that permit easy selection of transformed, transfected, transduced, or the like cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Suitable markers for yeast host cells include, but are not limited to, ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3.

The vectors may contain one or more (e.g., two, several) elements that permit integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome.

For integration into the host cell genome, the vector may rely on the polynucleotide's sequence encoding the polypeptide or any other element of the vector for integration into the genome by homologous or non-homologous recombination. Alternatively, the vector may contain additional polynucleotides for directing integration by homologous recombination into the genome of the host cell at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, 400 to 10,000 base pairs, and 800 to 10,000 base pairs, which have a high degree of sequence identity to the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding polynucleotides. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination. Potential integration loci include those described in the art (e.g., See US2012/0135481).

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the yeast cell. The origin of replication may be any plasmid replicator mediating autonomous replication that functions in a cell. The term "origin of replication" or "plasmid replicator" means a polynucleotide that enables a plasmid or vector to replicate *in vivo*. Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6.

More than one copy of a polynucleotide described herein may be inserted into a host cell to increase production of a polypeptide. An increase in the copy number of the polynucleotide can be obtained by integrating at least one additional copy of the sequence into the yeast cell genome

CA 03158982 2022-5-19

or by including an amplifiable selectable marker gene with the polynucleotide where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the polynucleotide, 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 described herein are well known to one skilled in the art (see, e.g., Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York).

Additional procedures and techniques known in the art for the preparation of recombinant cells for ethanol fermentation, are described in, e.g., WO 2016/045569, the content of which is hereby incorporated by reference.

The host cell or fermenting organism may be in the form of a composition comprising a host cell or fermenting organism (e.g., a yeast strain described herein) and a naturally occurring and/or a non-naturally occurring component.

The host cell or fermenting organism described herein may be in any viable form, including crumbled, dry, including active dry and instant, compressed, cream (liquid) form etc. In one embodiment, the host cell or fermenting organism (e.g., a *Saccharomyces cerevisiae* yeast strain) is dry yeast, such as active dry yeast or instant yeast. In one embodiment, the host cell or fermenting organism (e.g., a *Saccharomyces cerevisiae* yeast strain) is crumbled yeast. In one embodiment, the host cell or fermenting organism (e.g., a *Saccharomyces cerevisiae* yeast strain) is compressed yeast. In one embodiment, the host cell or fermenting organism (e.g., a *Saccharomyces cerevisiae* yeast strain) is cream yeast.

In one embodiment is a composition comprising a host cell or fermenting organism described herein (e.g., a *Saccharomyces cerevisiae* yeast strain), and one or more of the component selected from the group consisting of: surfactants, emulsifiers, gums, swelling agent, and antioxidants and other processing aids.

The compositions described herein may comprise a host cell or fermenting organism described herein (e.g., a *Saccharomyces cerevisiae* yeast strain) and any suitable surfactants. In one embodiment, the surfactant(s) is/are an anionic surfactant, cationic surfactant, and/or nonionic surfactant.

The compositions described herein may comprise a host cell or fermenting organism described herein (e.g., a *Saccharomyces cerevisiae* yeast strain) and any suitable emulsifier. In one embodiment, the emulsifier is a fatty-acid ester of sorbitan. In one embodiment, the emulsifier is selected from the group of sorbitan monostearate (SMS), citric acid esters of monodiglycerides, polyglycerolester, fatty acid esters of propylene glycol.

CA 03158982 2022-5-19

In one embodiment, the composition comprises a host cell or fermenting organism described herein (e.g., a *Saccharomyces cerevisiae* yeast strain), and Olindronal SMS, Olindronal SK, or Olindronal SPL including composition concerned in European Patent No. 1,724,336 (hereby incorporated by reference). These products are commercially available from Bussetti, Austria, for active dry yeast.

The compositions described herein may comprise a host cell or fermenting organism described herein (e.g., a *Saccharomyces cerevisiae* yeast strain) and any suitable gum. In one embodiment, the gum is selected from the group of carob, guar, tragacanth, arabic, xanthan and acacia gum, in particular for cream, compressed and dry yeast.

The compositions described herein may comprise a host cell or fermenting organism described herein (e.g., a *Saccharomyces cerevisiae* yeast strain) and any suitable swelling agent. In one embodiment, the swelling agent is methyl cellulose or carboxymethyl cellulose.

The compositions described herein may comprise a host cell or fermenting organism described herein (e.g., a *Saccharomyces cerevisiae* yeast strain) and any suitable anti-oxidant. In one embodiment, the antioxidant is butylated hydroxyanisol (BHA) and/or butylated hydroxytoluene (BHT), or ascorbic acid (vitamin C), particular for active dry yeast.

Sugar Transporters

In some embodiments, the fermenting organism (e.g., recombinant yeast cell) comprises a genetic modification that increases or decreases expression of a sugar transporter. The transporter may be any transporter that is suitable for improving the utilization of pentose of the fermenting organisms having an active pentose fermentation pathway, such as a naturally occurring transporter (e.g., a native transporter from another species or an endogenous transporter expressed from a modified expression vector) or a variant thereof that retains transporter activity.

Transporter activity can be measured using any suitable assay known in the art, such as improvements in overall consumption of and/or growth rate on arabinose and/or xylose as described in the Examples herein.

In some embodiments, the genetic modification is a heterologous polynucleotide encoding a sugar transporter.

In some embodiments, the fermenting organism has an increased level of transporter activity compared to the fermenting organism without the genetic modification, when cultivated under the same conditions. In some embodiments, the fermenting organism has an increased level of transporter activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least

CA 03158982 2022-5-19

25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the fermenting organism without the genetic modification, when cultivated under the same conditions.

In some embodiments, the fermenting organism has increased or decreased expression of a sugar transporter when compared to *Saccharomyces cerevisiae* strain Ethanol Red® (deposited under Accession No. V14/007039 at National Measurement Institute, Victoria, Australia) under the same conditions. In some embodiments, the fermenting organism has an increased expression of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to *Saccharomyces cerevisiae* strain Ethanol Red® (deposited under Accession No. V14/007039 at National Measurement Institute, Victoria, Australia) under the same conditions (e.g., under conditions described herein, such as on or after 53 hours fermentation).

Exemplary sugar transporters that may be expressed with the fermenting organisms and methods of use described herein include, but are not limited to, transporters shown in Table 1 (or derivatives thereof).

Table 1.

	Description	Gene accession	Gene source	Transporter SEQ ID NO.
1	Sucrose transport protein SUC2	A0A371ENF9	*Mucuna pruriens*	257
2	Putative Sugar transporter	EFP1CHF3L5	*Penicillium brevicompactum*	258
3	HXT2 (hexose transporter) protein	BFJ89633	*Saccharomyces cerevisiae*	259
4	gxf1p protein	BFJ94472	*Metschnikowia sp*	260
5	gxf2p protein	BFJ94474	*Metschnikowia sp*	261
6	GntR family transcriptional regulator	A0A2T0LIY3	*Planifilum fimeticola*	262
7	Putative Sugar transporter	EFP1C62R7C	*Candida boidinii*	263
8	Putative Sugar transporter	EFPBZZ6M7	*Zygosaccharomyces kombuchaensis*	264
9	Putative Sugar transporter	EFPBZZCJL	*Candida haemulonis*	265
10	Putative Sugar transporter	EFPBZD7P6	*Spathaspora sp.*	266
11	Putative Sugar transporter	EFPBZBQPC	*Metschnikowia fructicola*	267
12	Putative Sugar transporter	EFPB917WB	*Talaromyces adpressus*	268
13	MFS domain-containing protein	A0A2H0ZXG4	*Candida auris*	269
14	Putative Sugar transporter	EFP9SKDNC	*Penicillium tularense*	270
15	Putative Sugar transporter	EFP9CLGNG	*Meyerozyma caribbica*	271
16	GntR family transcriptional regulator	A0A235B4R8	*Paludifilum halophilum*	272
17	Putative Sugar transporter	EFP8ZW2C9	*Scheffersomyces stambukii*	273

CA 03158982 2022-5-19

18	Putative Sugar transporter	EFP8DZ469	*Torulaspora microellipsoides*	274
19	Putative Sugar transporter	EFP86V2P1	*Morchella semilibera*	275
20	Putative Sugar transporter	EFP7WSC34	*Spathaspora boniae*	276
21	Putative Sugar transporter	EFP7J7B0Q	*Ilyonectria destructans*	277
22	Putative Sugar transporter	EFP7HS9KT	*Sugiyamaella xylanicola*	278
23	Putative Sugar transporter	EFP7G7KNB	*Saccharomycopsis fibuligera*	279
24	Putative Sugar transporter	EFP7FXX0N	*Yarrowia alimentaria*	280
25	Putative Sugar transporter	EFP7FXX0V	*Yarrowia alimentaria*	281
26	Putative Sugar transporter	EFP7FXWPL	*Yarrowia galli*	282
27	Putative Sugar transporter	EFP7FXWQF	*Yarrowia galli*	283
28	Putative Sugar transporter	EFP7FXWHQ	*Yarrowia phangngaensis*	284
29	-------	A0A1G4JE77	*Lachancea meyersii*	285
30	-------	A0A1G4MBE8	*Lachancea fermentati*	286
31	-------	A0A1G4MC24	*Lachancea fermentati*	287
32	-------	A0A1E4T6F0	*Candida arabinofermentans*	288
33	Putative Sugar transporter	EFP713NV2	*Kluyveromyces marxianus*	289
34	Putative Sugar transporter	EFP6T73D9	*Debaryomyces hansenii*	290
35	Putative Sugar transporter	EFP6T7804	*Debaryomyces hansenii*	291
36	Putative Sugar transporter	EFP6T76PV	*Debaryomyces hansenii*	292
37	Putative Sugar transporter	EFP6RQ8JN	*Scheffersomyces stipitis*	293
38	Putative Sugar transporter	EFP6RN7NJ	*Schwanniomyces occidentalis*	294
39	Putative Sugar transporter	EFP6PD54N	*Wickerhamomyces anomalus*	295
40	Putative Sugar transporter	EFP6BNRRN	*Lachancea cidri*	296
41	Putative Sugar transporter	EFP6BNQR8	*Lachancea cidri*	297
42	Putative Sugar transporter	EFP5QXT3D	*Yarrowia deformans*	298
43	Putative Sugar transporter	EFP5QXVB6	*Yarrowia deformans*	299
44	Putative Sugar transporter	EFP5QZ5XB	*Yarrowia deformans*	300
45	Putative Sugar transporter	EFP5QNR84	*Ambrosiozyma monospora*	301
46	Putative Sugar transporter	EFP5Q5H1L	*Ogataea methanolica*	302
47	Putative Sugar transporter	EFP5NSWCS	*Candida succiphila*	303
48	Putative Sugar transporter	EFP5NR67S	*Candida carpophila*	304
49	Putative Sugar transporter	EFP5NRP7F	*Candida carpophila*	305
50	Putative Sugar transporter	EFP5NNT0L	*Wickerhamia fluorescens*	306
51	Putative Sugar transporter	EFP5N972X	*Priceomyces haplophilus*	307
52	High-affinity hexose transporter HXT6	A0A0W0DYZ4	*Candida glabrata*	308
53	Putative Sugar transporter	EFP5FS42L	*Candida sojae*	309
54	GntR family transcriptional regulator	A0A0U0CXJ1	*Streptococcus pneumoniae*	310
55	-------	A0A0P1KXV6	*Lachancea quebecensis*	311
56	-------	A0A0N7MLX3	*Lachancea quebecensis*	312
57	GXS1 protein mutant	BBZ79998	*Candida intermedia*	313
58	Putative Sugar transporter	EFP401CL1	*Penicillium vulpinum*	314
59	Putative Sugar transporter	EFP3TVZL9	*Ogataea methanolica*	315
60	Putative Sugar transporter	EFP3TTLXK	*Schwanniomyces occidentalis*	316

CA 03158982 2022-5-19

61	Putative Sugar transporter	EFP3FBKC6	*Kluyveromyces marxianus*	317
62	Putative Sugar transporter	EFP2C7LS6	*Aspergillus oryzae*	318
63	Putative Sugar transporter	EFP1D9FNG	*Spathaspora arborariae*	319
64	Putative Sugar transporter	EFP14W5DC	*Clavispora lusitaniae*	320
65	Putative Sugar transporter	EFP12DWXL	*Spathaspora passalidarum*	321
66	arabinose transporter (AraT)	BAO92398	*Penicillium chrysogenum*	322
67	Putative Sugar transporter	EFPN276J	*Kluyveromyces wickerhamii*	323
68	Putative Sugar transporter	EFPL4075	*Kluyveromyces aestuarii*	324
69	HXT10-like protein	J6ECW9	*Saccharomyces kudriavzevii*	325
70	HXT6-like protein	J4U468	*Saccharomyces kudriavzevii*	326
71	HXT2-like protein	J5S3S1	*Saccharomyces kudriavzevii*	327
72	Putative Sugar transporter	EFP4VJQD	*Trichophaea saccata*	328
73	MFS domain-containing protein	G8ZV29	*Torulaspora delbrueckii*	329
74	protein HGT2	G3AF26	*Spathaspora passalidarum*	330
75	GAL2p	D3XDC4	*Saccharomyces kudriavzevii*	331
76	Putative L-arabinose transporter	C8TEF4	*Candida arabinofermentans*	332
77	-------	C5DDE9	*Lachancea thermotolerans*	333
78	-------	C5DHA8	*Lachancea thermotolerans*	334
79	Putative Sugar transporter	ABN64726	*Scheffersomyces stipitis*	335
80	Putative Sugar transporter	CAG87483	*Debaryomyces hansenii*	336
81	Putative Sugar transporter	AAT95983	*Torulaspora delbrueckii*	337
82	Putative Sugar transporter	CAG60202	*Candida glabrata*	338
83	Putative Sugar transporter	CAG58441	*Candida glabrata*	339
84	Putative Sugar transporter	CAG57753	*Candida glabrata*	340
85	PgLAT2 protein	AOD63543	*Pichia guilliermondii*	341
86	Sucrose transport protein SUF1	A3DSX4	*Phaseolus vulgaris*	342
87	Sugar transport protein 2	Q9LNV3	*Arabidopsis thaliana*	343
88	-------	Q6CD11	*Yarrowia lipolytica*	344
89	Arabinose metabolism transcriptional repressor	P96711	*Bacillus subtilis*	345
90	High-affinity glucose transporter	P49374	*Kluyveromyces lactis*	346
91	Hexose transporter HXT10	P43581	*Saccharomyces cerevisiae*	347
92	High-affinity hexose transporter HXT7	P39004	*Saccharomyces cerevisiae*	348
93	Hexose transporter HXT9	P40885	*Saccharomyces cerevisiae*	349
94	Putative Sugar transporter	A0A090BHJ4	*Kluyveromyces marxianus*	350
95	Putative Sugar transporter	A0A1L0DZ32	*Candida intermedia*	351

CA 03158982 2022-5-19

96	Putative Sugar transporter	A0A0P1KWW5	*Lachancea quebecensis*	352
97	Putative Sugar transporter	A1DAC2	*Aspergillus fischeri*	353
98	Putative Sugar transporter	A0A1K2I4N4	*Lactobacillus rennini*	354
99	Putative Sugar transporter	D7KH13	*Arabidopsis lyrata*	355
100	Putative Sugar transporter	A0A1A0HK54	*Metschnikowia bicuspidata*	356
101	Putative Sugar transporter	A0A1S2Z5S7	*Cicer arietinum*	357
102	-------	A0A0L9VMD5	*Vigna angularis*	358
103	Sucrose transport protein	A0A1U9X406	*Pisum sativum*	359
104	Putative Sugar transporter	A0A1Q2ZT88	*Zygosaccharomyces rouxii*	360
105	Putative Sugar transporter	A0A0M0KUB1	*Jeotgalibacillus marinus*	361
106	Putative Sugar transporter	A0A202G714	*Clavispora lusitaniae*	362
107	Putative Sugar transporter	A0A202G702	*Clavispora lusitaniae*	363
108	Putative Sugar transporter	B6HE12	*Penicillium rubens*	364
109	Putative Sugar transporter	A0A0D4JCC0	*Saccharomyces cerevisiae*	365
110	-------	M5P6N0	*Bacillus sonorensis*	366
111	Putative Sugar transporter	A0A078DBU3	*Brassica napus*	367
112	Putative Sugar transporter	V4KEI8	*Eutrema salsugineum*	368
113	Putative Sugar transporter	A0A202G6Z7	*Clavispora lusitaniae*	369
114	Putative Sugar transporter	A0A1N6MBZ0	*Yarrowia galli*	370
115	Putative Sugar transporter	Q6BV56	*Debaryomyces hansenii*	371
116	Putative Sugar transporter	A5DPY9	*Meyerozyma guilliermondii*	372
117	Putative Sugar transporter	B1H0U7	*Ambrosiozyma monospora*	373
118	Putative Sugar transporter	B2G4F7	*Zygosaccharomyces rouxii*	374
119	Putative Sugar transporter	A0A1N6MBV7	*Yarrowia alimentaria*	375
120	Putative Sugar transporter	Q6CG69	*Yarrowia lipolytica*	376
121	Putative Sugar transporter	R0GW82	*Capsella rubella*	377
122	Putative Sugar transporter	A0A087HLR1	*Arabis alpina*	378
123	Putative Sugar transporter	B1H0U6	*Ambrosiozyma monospora*	379
124	Putative Sugar transporter	K9FYP3	*Penicillium digitatum*	380
125	Putative Sugar transporter	EFPC7NHF4	low complexity metagenome	381
126	Putative Sugar transporter	C4B4V9	*Corynebacterium glutamicum*	382
127	Putative Sugar transporter	A0A1L0BAU2	*Candida intermedia*	383
128	Putative Sugar transporter	A0A0R1SSI1	*Lactobacillus versmoldensis*	384
129	Putative Sugar transporter	A0A1L9U9S9	*Aspergillus brasiliensis*	385
130	Putative Sugar transporter	A5DWD7	*Lodderomyces elongisporus*	386
131	Putative Sugar transporter	A1C8W7	*Aspergillus clavatus*	387
132	Putative Sugar transporter	A0A1Y6JY60	*Lactobacillus zymae*	388
133	Putative Sugar transporter	A0A1G4MFR0	*Lachancea fermentati*	389
134	Putative Sugar transporter	A0A250WLN4	*Saccharomyces cerevisiae*	390
135	Putative Sugar transporter	K0KRN7	*Wickerhamomyces ciferrii*	391
136	Putative Sugar transporter	A0A1E4T2R0	*Candida arabinofermentans*	392

CA 03158982 2022-5-19

137	GntR family transcriptional regulator	A0A1Q9FY01	*Bacillus licheniformis*	393
138	Putative Sugar transporter	A0A1L0BZU1	*Candida intermedia*	394
139	Putative Sugar transporter	A0A0A8KZI3	*Kluyveromyces dobzhanskii*	395
140	Putative transporter	EFP70FSPD	*Saccharomyces cerevisiae*	396
141	Putative transporter	EFP7TC8PR	*Saccharomyces cerevisiae*	397

Additional polynucleotides encoding suitable sugar transporters may be derived from microorganisms of any suitable genus, including those readily available within the UniProtKB database (www.uniprot.org).

The sugar transporter may be a bacterial transporter. For example, the transporter may be derived from a Gram-positive bacterium such as a *Bacillus*, *Clostridium*, *Enterococcus*, *Geobacillus*, *Lactobacillus*, *Lactococcus*, *Oceanobacillus*, *Staphylococcus*, *Streptococcus*, or *Streptomyces*, or a Gram-negative bacterium such as a *Campylobacter*, *E. coli*, *Flavobacterium*, *Fusobacterium*, *Helicobacter*, *Ilyobacter*, *Neisseria*, *Pseudomonas*, *Salmonella*, or *Ureaplasma*.

In one embodiment, the sugar transporter is derived from *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, or *Bacillus thuringiensis*.

In another embodiment, the sugar transporter is derived from *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, or *Streptococcus equi* subsp. *Zooepidemicus*.

In another embodiment, the sugar transporter is derived from *Streptomyces achromogenes*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, or *Streptomyces lividans*.

The sugar transporter may be a fungal transporter. For example, the sugar transporter may be derived from a yeast such as a *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, *Yarrowia* or *Issatchenkia*; or derived from a filamentous fungus such as an *Acremonium*, *Agaricus*, *Alternaria*, *Aspergillus*, *Aureobasidium*, *Botryospaeria*, *Ceriporiopsis*, *Chaetomidium*, *Chrysosporium*, *Claviceps*, *Cochliobolus*, *Coprinopsis*, *Coptotermes*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Diplodia*, *Exidia*, *Filibasidium*, *Fusarium*, *Gibberella*, *Holomastigotoides*, *Humicola*, *Irpex*, *Lentinula*, *Leptospaeria*, *Magnaporthe*, *Melanocarpus*, *Meripilus*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Piromyces*, *Poitrasia*, *Pseudoplectania*, *Pseudotrichonympha*, *Rhizomucor*,

CA 03158982 2022-5-19

Schizophyllum, *Scytalidium*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trichoderma*, *Trichophaea*, *Verticillium*, *Volvariella*, or *Xylaria*.

In another embodiment, the sugar transporter is derived from *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, or *Saccharomyces oviformis*.

In another embodiment, the sugar transporter is derived from *Acremonium cellulolyticus*, *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus fumigatus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Chrysosporium inops*, *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium merdarium*, *Chrysosporium pannicola*, *Chrysosporium queenslandicum*, *Chrysosporium tropicum*, *Chrysosporium zonatum*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola grisea*, *Humicola insolens*, *Humicola lanuginosa*, *Irpex lacteus*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium funiculosum*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Thielavia achromatica*, *Thielavia albomyces*, *Thielavia albopilosa*, *Thielavia australeinsis*, *Thielavia fimeti*, *Thielavia microspora*, *Thielavia ovispora*, *Thielavia peruviana*, *Thielavia setosa*, *Thielavia spededonium*, *Thielavia subthermophila*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, or *Trichoderma viride*.

It will be understood that for the aforementioned species, the invention encompasses both the perfect and imperfect states, and other taxonomic equivalents, e.g., anamorphs, regardless of the species name by which they are known. Those skilled in the art will readily recognize the identity of appropriate equivalents.

Strains of these species are readily accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

The sugar transporter coding sequences described or referenced herein, or a subsequence thereof, as well as the transporter described or referenced herein, or a fragment thereof, may be used to design nucleic acid probes to identify and clone DNA encoding a

CA 03158982 2022-5-19

transporter from strains of different genera or species according to methods well known in the art. In particular, such probes can be used for hybridization with the genomic DNA or cDNA of a cell of interest, following standard Southern blotting procedures, in order to identify and isolate the corresponding gene therein. Such probes can be considerably shorter than the entire sequence, but should be at least 15, e.g., at least 25, at least 35, or at least 70 nucleotides in length. Preferably, the nucleic acid probe is at least 100 nucleotides in length, e.g., at least 200 nucleotides, at least 300 nucleotides, at least 400 nucleotides, at least 500 nucleotides, at least 600 nucleotides, at least 700 nucleotides, at least 800 nucleotides, or at least 900 nucleotides in length. Both DNA and RNA probes can be used. The probes are typically labeled for detecting the corresponding gene (for example, with ^{32}P, ^{3}H, ^{35}S, biotin, or avidin).

A genomic DNA or cDNA library prepared from such other strains may be screened for DNA that hybridizes with the probes described above and encodes a sugar transporter. Genomic or other DNA from such other strains may be separated by agarose or polyacrylamide gel electrophoresis, or other separation techniques. DNA from the libraries or the separated DNA may be transferred to and immobilized on nitrocellulose or other suitable carrier material. In order to identify a clone or DNA that hybridizes with a coding sequence, or a subsequence thereof, the carrier material is used in a Southern blot.

In one embodiment, the nucleic acid probe is a polynucleotide, or subsequence thereof, that encodes the sugar transporter of any one of SEQ ID NOs: 257-397, or a fragment thereof.

For purposes of the probes described above, hybridization indicates that the polynucleotide hybridizes to a labeled nucleic acid probe, or the full-length complementary strand thereof, or a subsequence of the foregoing; under very low to very high stringency conditions. Molecules to which the nucleic acid probe hybridizes under these conditions can be detected using, for example, X-ray film. Stringency and washing conditions are defined as described *supra*.

In one embodiment, the sugar transporter is encoded by a polynucleotide that hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence for any one of the transporters described or referenced herein (e.g., SEQ ID NOs: 257-397). (Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York).

The sugar transporter may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, silage, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, silage, etc.) using the above-mentioned probes. Techniques for isolating microorganisms and DNA directly from natural

CA 03158982 2022-5-19

habitats are well known in the art. The polynucleotide encoding a sugar transporter may then be derived by similarly screening a genomic or cDNA library of another microorganism or mixed DNA sample.

Once a polynucleotide encoding a sugar transporter has been detected with a suitable probe as described herein, the sequence may be isolated or cloned by utilizing techniques that are known to those of ordinary skill in the art (*See*, e.g., Sambrook et al., 1989, *supra*). Techniques used to isolate or clone polynucleotides encoding transporters include isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the polynucleotides from such genomic DNA can be affected, e.g., by using the well-known polymerase chain reaction (PCR) or antibody screening of expression libraries to detect cloned DNA fragments with shares structural features (*See*, e.g., Innis et al., 1990, *PCR: A Guide to Methods and Application*, Academic Press, New York). Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleotide sequence-based amplification (NASBA) may be used.

In one embodiment, the sugar transporter comprises or consists of the amino acid sequence of any one of SEQ ID NOs: 257-397 (such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138). In another embodiment, the transporter is a fragment of the transporter of any one of SEQ ID NOs: 257-397 (such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138), wherein, e.g., the fragment has transporter activity. In one embodiment, the number of amino acid residues in the fragment is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of amino acid residues in referenced full length transporter (e.g. any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138)). In other embodiments, the transporter may comprise the catalytic domain of any transporter described or referenced herein (e.g., the catalytic domain of any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138).

The transporter may be a variant of any one of the transporter described *supra* (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138). In one embodiment, the transporter has at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of the transporters described *supra* (e.g., any one of

CA 03158982 2022-5-19

SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138).

In one embodiment, the transporter sequence differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from the amino acid sequence of any one of the transportera described *supra* (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138). In one embodiment, the transporter has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) of amino acid sequence of any one of the transporters described *supra* (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138). In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the sugar transporter is not the transporter having a mature polypeptide sequence of SEQ ID NO: 390 (or a transporter having a mature polypeptide sequence with at least 80%, e.g., at least 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to the transporter of SEQ ID NO: 390).

The amino acid changes are generally of a minor nature, that is conservative amino acid substitutions or insertions that do not significantly affect the folding and/or activity of the protein; small deletions, typically of one to about 30 amino acids; small amino-terminal or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to about 20-25 residues; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

Examples of conservative substitutions are within the group of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine and methionine). Amino acid substitutions that do not generally alter specific activity are known in the art and are described, for example, by H. Neurath and R.L. Hill, 1979, *In*, *The Proteins*, Academic Press, New York. The most commonly occurring exchanges are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly.

Alternatively, the amino acid changes are of such a nature that the physico-chemical properties of the polypeptides are altered. For example, amino acid changes may improve the

CA 03158982 2022-5-19

thermal stability of the transporters, alter the substrate specificity, change the pH optimum, and the like.

Essential amino acids can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, 1989, *Science* 244: 1081-1085). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant mutant molecules are tested for activity to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton *et al.*, 1996, *J. Biol. Chem.* 271: 4699-4708. The active site or other biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction, or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids (*See*, for example, de Vos *et al.*, 1992, *Science* 255: 306-312; Smith *et al.*, 1992, *J. Mol. Biol.* 224: 899-904; Wlodaver *et al.*, 1992, *FEBS Lett.* 309: 59-64). The identities of essential amino acids can also be inferred from analysis of identities with other sugar transporters that are related to the referenced transporter.

Additional guidance on the structure-activity relationship of the transporters herein can be determined using multiple sequence alignment (MSA) techniques well-known in the art. Based on the teachings herein, the skilled artisan could make similar alignments with any number of transporters described herein or known in the art. Such alignments aid the skilled artisan to determine potentially relevant domains (e.g., binding domains or catalytic domains), as well as which amino acid residues are conserved and not conserved among the different transporter sequences. It is appreciated in the art that changing an amino acid that is conserved at a particular position between disclosed polypeptides will more likely result in a change in biological activity (Bowie *et al.*, 1990, *Science* 247: 1306-1310: "Residues that are directly involved in protein functions such as binding or catalysis will certainly be among the most conserved"). In contrast, substituting an amino acid that is not highly conserved among the polypeptides will not likely or significantly alter the biological activity.

Even further guidance on the structure-activity relationship for the skilled artisan can be found in published x-ray crystallography studies known in the art. As noted *supra*, additional characterization of AAAPs are described, e.g., Young et al., 1999, *Biochimica et Biophysica Acta* 1415: 306-322. Structure-function analysis is also described, e.g., Swarup et al., 2004, *The Plant Cell*, 16:3069-3083.

Single or multiple amino acid substitutions, deletions, and/or insertions can be made and tested using known methods of mutagenesis, recombination, and/or shuffling, followed by a relevant screening procedure, such as those disclosed by Reidhaar-Olson and Sauer, 1988,

CA 03158982 2022-5-19

Science 241: 53-57; Bowie and Sauer, 1989, *Proc. Natl. Acad. Sci. USA* 86: 2152-2156; WO95/17413; or WO95/22625. Other methods that can be used include error-prone PCR, phage display (e.g., Lowman *et al.*, 1991, *Biochemistry* 30: 10832-10837; U.S. Patent No. 5,223,409; WO92/06204), and region-directed mutagenesis (Derbyshire *et al.*, 1986, *Gene* 46: 145; Ner *et al.*, 1988, *DNA* 7: 127).

Mutagenesis/shuffling methods can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides expressed by host cells (Ness *et al.*, 1999, *Nature Biotechnology* 17: 893-896). Mutagenized DNA molecules that encode active transporters can be recovered from the host cells and rapidly sequenced using standard methods in the art. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide.

In another embodiment, the heterologous polynucleotide encoding the sugar transporter comprises a coding sequence having at least 60%, e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the coding sequence of any one of the transporters described *supra* (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138).

In one embodiment, the heterologous polynucleotide encoding the sugar transporter comprises or consists of the coding sequence of any one of the transporters described *supra* (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138). In another embodiment, the heterologous polynucleotide encoding the sugar transporter comprises a subsequence of the coding sequence of any one of the transporters described *supra* (e.g., any one of SEQ ID NOs: 257-397; such as any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131; and/or any one of SED ID NOs: 97, 116 and 138) wherein the subsequence encodes a polypeptide having transporter activity. In another embodiment, the number of nucleotides residues in the coding subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The referenced coding sequence of any related aspect or embodiment described herein can be the native coding sequence or a degenerate sequence, such as a codon-optimized coding sequence designed for use in a particular host cell (e.g., optimized for expression in *Saccharomyces cerevisiae*).

The sugar transporter may be a fused polypeptide or cleavable fusion polypeptide in which another polypeptide is fused at the N-terminus or the C-terminus of the transporter. A fused

polypeptide may be produced by fusing a polynucleotide encoding another polypeptide to a polynucleotide encoding the transporter. Techniques for producing fusion polypeptides are known in the art, and include ligating the coding sequences encoding the polypeptides so that they are in frame and that expression of the fused polypeptide is under control of the same promoter(s) and terminator. Fusion proteins may also be constructed using intein technology in which fusions are created post-translationally (Cooper *et al.*, 1993, *EMBO J.* 12: 2575-2583; Dawson *et al.*, 1994, *Science* 266: 776-779).

In some embodiments, the sugar transporter is a fusion protein comprising a signal peptide linked to the N-terminus of a mature polypeptide, such as any signal sequences described in U.S. Provisional Application No. 62/883,519 filed August 6, 2019 and entitled "Fusion Proteins For Improved Enzyme Expression" (the content of which is hereby incorporated by reference).

Non-phosphorylating NADP-dependent glyceraldehyde-3-phosphate dehydrogenases (GAPNs)

The host cells and fermenting organisms may express a heterologous NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN). The GAPN can be any GAPN that is suitable for the host cells and their methods of use described herein, such as a naturally occurring GAPN (e.g., an endogenous GAPN or a native GAPN from another species) or a variant thereof that retains GAPN activity. In one aspect, GAPN is present in the cytosol of the host cells.

In some embodiments, the host cell or fermenting organism comprises a heterologous polynucleotide encoding a GAPN. In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding a GAPN has an increased level of GAPN activity compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the GAPN, when cultivated under the same conditions. In some embodiments, the host cell or fermenting organism has an increased level of GAPN activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to host cell or fermenting organism without the heterologous polynucleotide encoding the GAPN, when cultivated under the same conditions.

Exemplary GAPNs that may be expressed with the host cells or fermenting organisms and methods of use described herein include, but are not limited to, GAPNs shown in Table 2 (or derivatives thereof).

Table 2.

	Donor Organism	Sequence code	SEQ ID NO.
1	*Triticum aestivum*	Q8LK61	407
2	*Chlamydomonas reinhardtii*	A0A2K3D5S6	408

CA 03158982 2022-5-19

3	*Apium graveolens*	Q9SNX8	409
4	*Cicer arietinum*	A0A1S2YP36	410
5	*Bacillus pseudomycoides*	A0A2C4I5G8	411
6	*Streptococcus equinus*	Q3C1A6	412
7	*Glycine soja*	A0A0B2QEZ3	413
8	*Streptococcus sp. DD12*	A0A139NKR4	414
9	*Bacillus thuringiensis*	A0A0B5NZK7	415
10	*Arabidopsis thaliana*	Q1WIQ6	416
11	*Bacillus litoralis*	EFP8C9GVR	417
12	*Streptococcus hyointestinalis*	A0A380K8A8	418
13	*Zea mays*	Q43272	419
14	*Lactobacillus delbrueckii*	Q04A83	420
15	*Streptococcus pluranimalium*	A0A2L0D390	421
16	*Nicotiana plumbaginifolia*	P93338	422
17	*Streptococcus macacae*	G5JUQ8	423
18	*Streptococcus mutans*	Q59931	424
19	*Bacillus cereus*		425
20	*Streptococcus thermophilus*		426
21	*Streptococcus urinalis*		427
22	*Streptococcus canis*		428
23	*Streptococcus thoraltensis*		429
24	*Streptococcus dysgalactiae*		430
25	*Streptococcus pyogenes*		431
26	*Streptococcus ictaluri*		432
27	*Clostridium perfringens*		433
28	*Clostridium chromiireducens*		434
29	*Clostridium botulinum*		435
30	*Bacillus anthracis*		436
31	*Pyrococcus furiosus*		437

Additional polynucleotides encoding suitable GAPNs may be derived from microorganisms of any suitable genus, including those readily available within the UniProtKB database (www.uniprot.org).

The GAPN coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding proteases from strains of different genera or species, as described *supra*.

The polynucleotides encoding GAPNs may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding GAPNs are described *supra*.

In one embodiment, the GAPN has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any one of SEQ ID NOs: 407-437. In another embodiment, the GAPN has a mature polypeptide sequence that is a fragment of the GAPN of any one of SEQ

CA 03158982 2022-5-19

ID NOs: 407-437 (e.g., wherein the fragment has GAPN activity). In one embodiment, the number of amino acid residues in the fragment is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of amino acid residues in referenced full length GAPN (e.g. any one of SEQ ID NOs: 407-437). In other embodiments, the GAPN may comprise the catalytic domain of any GAPN described or referenced herein (e.g., the catalytic domain of any one of SEQ ID NOs: 407-437).

The GAPN may be a variant of any one of the GAPNs described *supra* (e.g., any one of SEQ ID NOs: 407-437). In one embodiment, the GAPN has a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of the GAPNs described *supra* (e.g., any one of SEQ ID NOs: 407-437).

Examples of suitable amino acid changes, such as conservative substitutions that do not significantly affect the folding and/or activity of the GAPN, are described *supra*.

In one embodiment, the GAPN has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from the amino acid sequence of any one of the GAPN described *supra* (e.g., any one of SEQ ID NOs: 407-437). In one embodiment, the GAPN has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) of amino acid sequence of any one of the GAPNs described *supra* (e.g., any one of SEQ ID NOs: 407-437). In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In one embodiment, the GAPN coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any GAPN described or referenced herein (e.g., any one of SEQ ID NOs: 407-437). In one embodiment, the GAPN coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any GAPN described or referenced herein (e.g., any one of SEQ ID NOs: 407-437).

In one embodiment, the GAPN comprises the coding sequence of any GAPN described or referenced herein (any one of SEQ ID NOs: 407-437). In one embodiment, the GAPN comprises a coding sequence that is a subsequence of the coding sequence from any GAPN described or referenced herein, wherein the subsequence encodes a polypeptide having GAPN

CA 03158982 2022-5-19

activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The referenced GAPN coding sequence of any related aspect or embodiment described herein can be the native coding sequence or a degenerate sequence, such as a codon-optimized coding sequence designed for use in a particular host cell (e.g., optimized for expression in *Saccharomyces cerevisiae*).

The GAPN can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

Glucoamylases

The host cells and fermenting organisms may express a heterologous glucoamylase. The glucoamylase can be any glucoamylase that is suitable for the host cells, fermenting organisms and/or their methods of use described herein, such as a naturally occurring glucoamylase or a variant thereof that retains glucoamylase activity. Any glucoamylase contemplated for expression by a host cell or fermenting organism described below is also contemplated for embodiments of the invention involving exogenous addition of a glucoamylase (e.g., added before, during or after liquefaction and/or saccharification).

In some embodiments, the host cell or fermenting organism comprises a heterologous polynucleotide encoding a glucoamylase, for example, as described in WO2017/087330, the content of which is hereby incorporated by reference. Any glucoamylase described or referenced herein is contemplated for expression in the host cell or fermenting organism.

In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding a glucoamylase has an increased level of glucoamylase activity compared to the host cells without the heterologous polynucleotide encoding the glucoamylase, when cultivated under the same conditions. In some embodiments, the host cell or fermenting organism has an increased level of glucoamylase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the glucoamylase, when cultivated under the same conditions.

Exemplary glucoamylases that can be used with the host cells and/or the methods described herein include bacterial, yeast, or filamentous fungal glucoamylases, e.g., obtained from any of the microorganisms described or referenced herein, as described *supra*.

CA 03158982 2022-5-19

Preferred glucoamylases are of fungal or bacterial origin, selected from the group consisting of *Aspergillus* glucoamylases, in particular *Aspergillus niger* G1 or G2 glucoamylase (Boel et al. (1984), EMBO J. 3 (5), p. 1097-1102), or variants thereof, such as those disclosed in WO 92/00381, WO 00/04136 and WO 01/04273 (from Novozymes, Denmark); the *A. awamori* glucoamylase disclosed in WO 84/02921, *Aspergillus oryzae* glucoamylase (Agric. Biol. Chem. (1991), 55 (4), p. 941-949), or variants or fragments thereof. Other *Aspergillus* glucoamylase variants include variants with enhanced thermal stability: G137A and G139A (Chen et al. (1996), Prot. Eng. 9, 499-505); D257E and D293E/Q (Chen et al. (1995), Prot. Eng. 8, 575-582); N182 (Chen et al. (1994), Biochem. J. 301, 275-281); disulphide bonds, A246C (Fierobe et al. (1996), Biochemistry, 35, 8698-8704; and introduction of Pro residues in position A435 and S436 (Li et al. (1997), Protein Eng. 10, 1199-1204.

Other glucoamylases include *Athelia rolfsii* (previously denoted *Corticium rolfsii*) glucoamylase (see US patent no. 4,727,026 and (Nagasaka et al. (1998) "Purification and properties of the raw-starch-degrading glucoamylases from *Corticium rolfsii*, Appl Microbiol Biotechnol 50:323-330), *Talaromyces* glucoamylases, in particular derived from *Talaromyces emersonii* (WO 99/28448), *Talaromyces leycettanus* (US patent no. Re. 32,153), *Talaromyces duponti, Talaromyces thermophilus* (US patent no. 4,587,215). In one embodiment, the glucoamylase used during saccharification and/or fermentation is the *Talaromyces emersonii* glucoamylase disclosed in WO 99/28448 or the *Talaromyces emersonii* glucoamylase of SEQ ID NO: 247.

Bacterial glucoamylases contemplated include glucoamylases from the genus *Clostridium*, in particular *C. thermoamylolyticum* (EP 135,138), and *C. thermohydrosulfuricum* (WO 86/01831).

Contemplated fungal glucoamylases include *Trametes cingulate*, *Pachykytospora papyracea*; and *Leucopaxillus giganteus* all disclosed in WO 2006/069289; or *Peniophora rufomarginata* disclosed in WO2007/124285; or a mixture thereof. Also hybrid glucoamylase are contemplated. Examples include the hybrid glucoamylases disclosed in WO 2005/045018.

In one embodiment, the glucoamylase is derived from a strain of the genus *Pycnoporus*, in particular a strain of *Pycnoporus* as described in WO 2011/066576 (SEQ ID NO: 2, 4 or 6 therein), including the *Pycnoporus sanguineus* glucoamylase, or from a strain of the genus *Gloeophyllum*, such as a strain of *Gloeophyllum sepiarium* or *Gloeophyllum trabeum*, in particular a strain of *Gloeophyllum* as described in WO 2011/068803 (SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16 therein). In one embodiment, the glucoamylase is SEQ ID NO: 2 in WO 2011/068803 (i.e. *Gloeophyllum sepiarium* glucoamylase). In one embodiment, the glucoamylase is the

CA 03158982 2022-5-19

Gloeophyllum sepiarium glucoamylase of SEQ ID NO: 8. In one embodiment, the glucoamylase is the *Pycnoporus sanguineus* glucoamylase of SEQ ID NO: 229.

In one embodiment, the glucoamylase is a *Gloeophyllum trabeum* glucoamylase (disclosed as SEQ ID NO: 3 in WO2014/177546). In another embodiment, the glucoamylase is derived from a strain of the genus *Nigrofomes,* in particular a strain of *Nigrofomes sp.* disclosed in WO 2012/064351 (disclosed as SEQ ID NO: 2 therein).

Also contemplated are glucoamylases with a mature polypeptide sequence which exhibit a high identity to any of the above mentioned glucoamylases, i.e., at least 60%, such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or even 100% identity to any one of the mature polypeptide sequences mentioned above.

Glucoamylases may be added to the saccharification and/or fermentation in an amount of 0.0001-20 AGU/g DS, such as 0.001-10 AGU/g DS, 0.01-5 AGU/g DS, or 0.1-2 AGU/g DS.

Glucoamylases may be added to the saccharification and/or fermentation in an amount of 1-1,000 µg EP/g DS, such as 10-500 µg/gDS, or 25-250 µg/g DS.

Glucoamylases may be added to liquefaction in an amount of 0.1-100 µg EP/g DS, such as 0.5-50 µg EP/g DS, 1-25 µg EP/g DS, or 2-12 µg EP/g DS.

In one embodiment, the glucoamylase is added as a blend further comprising an alpha-amylase (e.g., any alpha-amylase described herein). In one embodiment, the alpha-amylase is a fungal alpha-amylase, especially an acid fungal alpha-amylase. The alpha-amylase is typically a side activity.

In one embodiment, the glucoamylase is a blend comprising *Talaromyces emersonii* glucoamylase disclosed in WO 99/28448 as SEQ ID NO: 34 and *Trametes cingulata* glucoamylase disclosed as SEQ ID NO: 2 in WO 06/069289.

In one embodiment, the glucoamylase is a blend comprising *Talaromyces emersonii* glucoamylase disclosed in WO 99/28448, *Trametes cingulata* glucoamylase disclosed as SEQ ID NO: 2 in WO 06/69289, and an alpha-amylase.

In one embodiment, the glucoamylase is a blend comprising *Talaromyces emersonii* glucoamylase disclosed in WO99/28448, *Trametes cingulata* glucoamylase disclosed in WO 06/69289, and *Rhizomucor pusillus* alpha-amylase with *Aspergillus niger* glucoamylase linker and SBD disclosed as V039 in Table 5 in WO 2006/069290.

In one embodiment, the glucoamylase is a blend comprising *Gloeophyllum sepiarium* glucoamylase shown as SEQ ID NO: 2 in WO 2011/068803 and an alpha-amylase, in particular *Rhizomucor pusillus* alpha-amylase with an *Aspergillus niger* glucoamylase linker and starch-

CA 03158982 2022-5-19

binding domain (SBD), disclosed SEQ ID NO: 3 in WO 2013/006756, in particular with the following substitutions: G128D+D143N.

In one embodiment, the alpha-amylase may be derived from a strain of the genus *Rhizomucor*, preferably a strain the *Rhizomucor pusillus*, such as the one shown in SEQ ID NO: 3 in WO2013/006756, or the genus *Meripilus*, preferably a strain of *Meripilus giganteus*. In one embodiment, the alpha-amylase is derived from a *Rhizomucor pusillus* with an *Aspergillus niger* glucoamylase linker and starch-binding domain (SBD), disclosed as V039 in Table 5 in WO 2006/069290.

In one embodiment, the *Rhizomucor pusillus* alpha-amylase or the *Rhizomucor pusillus* alpha-amylase with an *Aspergillus niger* glucoamylase linker and starch-binding domain (SBD) has at least one of the following substitutions or combinations of substitutions: D165M; Y141W; Y141R; K136F; K192R; P224A; P224R; S123H+Y141W; G20S + Y141W; A76G + Y141W; G128D + Y141W; G128D + D143N; P219C + Y141W; N142D + D143N; Y141W + K192R; Y141W + D143N; Y141W + N383R; Y141W + P219C + A265C; Y141W + N142D + D143N; Y141W + K192R V410A; G128D + Y141W + D143N; Y141W + D143N + P219C; Y141W + D143N + K192R; G128D + D143N + K192R; Y141W + D143N + K192R + P219C; and G128D + Y141W + D143N + K192R; or G128D + Y141W + D143N + K192R + P219C (using SEQ ID NO: 3 in WO 2013/006756 for numbering).

In one embodiment, the glucoamylase blend comprises *Gloeophyllum sepiarium glucoamylase* (e.g., SEQ ID NO: 2 in WO 2011/068803) and *Rhizomucor pusillus* alpha-amylase.

In one embodiment, the glucoamylase blend comprises *Gloeophyllum sepiarium* glucoamylase shown as SEQ ID NO: 2 in WO 2011/068803 and *Rhizomucor pusillus* with an *Aspergillus niger* glucoamylase linker and starch-binding domain (SBD), disclosed SEQ ID NO: 3 in WO 2013/006756 with the following substitutions: G128D+D143N.

Commercially available compositions comprising glucoamylase include AMG 200L; AMG 300 L; SAN™ SUPER, SAN™ EXTRA L, SPIRIZYME® PLUS, SPIRIZYME® FUEL, SPIRIZYME® B4U, SPIRIZYME® ULTRA, SPIRIZYME® EXCEL, SPIRIZYME ACHIEVE®, and AMG® E (from Novozymes A/S); OPTIDEX™ 300, GC480, GC417 (from DuPont-Danisco); AMIGASE™ and AMIGASE™ PLUS (from DSM); G-ZYME™ G900, G-ZYME™ and G990 ZR (from DuPont-Danisco).

In one embodiment, the glucoamylase is derived from the *Debaryomyces occidentalis* glucoamylase of SEQ ID NO: 102. In one embodiment, the glucoamylase is derived from the *Saccharomycopsis fibuligera* glucoamylase of SEQ ID NO: 103. In one embodiment, the glucoamylase is derived from the *Saccharomycopsis fibuligera* glucoamylase of SEQ ID NO: 104.

CA 03158982 2022-5-19

In one embodiment, the glucoamylase is derived from the *Saccharomyces cerevisiae* glucoamylase of SEQ ID NO: 105. In one embodiment, the glucoamylase is derived from the *Aspergillus niger* glucoamylase of SEQ ID NO: 106. In one embodiment, the glucoamylase is derived from the *Aspergillus oryzae* glucoamylase of SEQ ID NO: 107. In one embodiment, the glucoamylase is derived from the *Rhizopus oryzae* glucoamylase of SEQ ID NO: 108 or SEQ ID NO: 250. In one embodiment, the glucoamylase is derived from the *Clostridium thermocellum* glucoamylase of SEQ ID NO: 109. In one embodiment, the glucoamylase is derived from the *Clostridium thermocellum* glucoamylase of SEQ ID NO: 110. In one embodiment, the glucoamylase is derived from the *Arxula adeninivorans* glucoamylase of SEQ ID NO: 111. In one embodiment, the glucoamylase is derived from the *Hormoconis resinae* glucoamylase of SEQ ID NO: 112. In one embodiment, the glucoamylase is derived from the *Aureobasidium pullulans* glucoamylase of SEQ ID NO: 113. In one embodiment, the glucoamylase is derived from the *Rhizopus microsporus* glucoamylase of SEQ ID NO: 248. In one embodiment, the glucoamylase is derived from the *Rhizopus delemar* glucoamylase of SEQ ID NO: 249. In one embodiment, the glucoamylase is derived from the *Punctularia strigosozonata* glucoamylase of SEQ ID NO: 244. In one embodiment, the glucoamylase is derived from the *Fibroporia radiculosa* glucoamylase of SEQ ID NO: 245. In one embodiment, the glucoamylase is derived from the *Wolfiporia cocos* glucoamylase of SEQ ID NO: 246.

In one embodiment, the glucoamylase is a *Trichoderma reesei* glucoamylase, such as the*Trichoderma reesei* glucoamylase of SEQ ID NO: 230.

In one embodiment, the glucoamylase has a Relative Activity heat stability at 85ºC of at least 20%, at least 30%, or at least 35% determined as described in Example 4 of WO2018/098381 (heat stability).

In one embodiment, the glucoamylase has a relative activity pH optimum at pH 5.0 of at least 90%, e.g., at least 95%, at least 97%, or 100% determined as described in Example 4 of WO2018/098381 (pH optimum).

In one embodiment, the glucoamylase has a pH stability at pH 5.0 of at least 80%, at least 85%, at least 90% determined as described in Example 4 of WO2018/098381 (pH stability).

In one embodiment, the glucoamylase used in liquefaction, such as a *Penicillium oxalicum* glucoamylase variant, has a thermostability determined as DSC Td at pH 4.0 as described in Example 15 of WO2018/098381 of at least 70ºC, preferably at least 75ºC, such as at least 80ºC, such as at least 81ºC, such as at least 82ºC, such as at least 83ºC, such as at least 84ºC, such as at least 85ºC, such as at least 86ºC, such as at least 87%, such as at least 88ºC, such as at least 89ºC, such as at least 90ºC. In one embodiment, the glucoamylase, such as a *Penicillium*

CA 03158982 2022-5-19

oxalicum glucoamylase variant, has a thermostability determined as DSC Td at pH 4.0 as described in Example 15 of WO2018/098381 in the range between 70°C and 95°C, such as between 80°C and 90°C.

In one embodiment, the glucoamylase, such as a *Penicillium oxalicum* glucoamylase variant, used in liquefaction has a thermostability determined as DSC Td at pH 4.8 as described in Example 15 of WO2018/098381 of at least 70°C, preferably at least 75°C, such as at least 80°C, such as at least 81°C, such as at least 82°C, such as at least 83°C, such as at least 84°C, such as at least 85°C, such as at least 86°C, such as at least 87%, such as at least 88°C, such as at least 89°C, such as at least 90°C, such as at least 91°C. In one embodiment, the glucoamylase, such as a *Penicillium oxalicum* glucoamylase variant, has a thermostability determined as DSC Td at pH 4.8 as described in Example 15 of WO2018/098381 in the range between 70°C and 95°C, such as between 80°C and 90°C.

In one embodiment, the glucoamylase, such as a *Penicillium oxalicum* glucoamylase variant, used in liquefaction has a residual activity determined as described in Example 16 of WO2018/098381, of at least 100% such as at least 105%, such as at least 110%, such as at least 115%, such as at least 120%, such as at least 125%. In one embodiment, the glucoamylase, such as a *Penicillium oxalicum* glucoamylase variant, has a thermostability determined as residual activity as described in Example 16 of WO2018/098381, in the range between 100% and 130%.

In one embodiment, the glucoamylase, e.g., of fungal origin such as a filamentous fungi, from a strain of the genus *Penicillium*, e.g., a strain of *Penicillium oxalicum*, in particular the *Penicillium oxalicum* glucoamylase disclosed as SEQ ID NO: 2 in WO 2011/127802 (which is hereby incorporated by reference).

In one embodiment, the glucoamylase has a mature polypeptide sequence of at least 80%, e.g., at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to the mature polypeptide shown in SEQ ID NO: 2 in WO 2011/127802.

In one embodiment, the glucoamylase is a variant of the *Penicillium oxalicum* glucoamylase disclosed as SEQ ID NO: 2 in WO 2011/127802, having a K79V substitution. The K79V glucoamylase variant has reduced sensitivity to protease degradation relative to the parent as disclosed in WO 2013/036526 (which is hereby incorporated by reference).

In one embodiment, the glucoamylase is derived from *Penicillium oxalicum*.

In one embodiment, the glucoamylase is a variant of the *Penicillium oxalicum* glucoamylase disclosed as SEQ ID NO: 2 in WO 2011/127802. In one embodiment, the

CA 03158982 2022-5-19

Penicillium oxalicum glucoamylase is the one disclosed as SEQ ID NO: 2 in WO 2011/127802 having Val (V) in position 79.

Contemplated *Penicillium oxalicum* glucoamylase variants are disclosed in WO 2013/053801 which is hereby incorporated by reference.

In one embodiment, these variants have reduced sensitivity to protease degradation.

In one embodiment, these variants have improved thermostability compared to the parent.

In one embodiment, the glucoamylase has a K79V substitution (using SEQ ID NO: 2 of WO 2011/127802 for numbering), corresponding to the PE001 variant, and further comprises one of the following alterations or combinations of alterations

T65A; Q327F; E501V; Y504T; Y504*; T65A + Q327F; T65A + E501V; T65A + Y504T; T65A + Y504*; Q327F + E501V; Q327F + Y504T; Q327F + Y504*; E501V + Y504T; E501V + Y504*; T65A + Q327F + E501V; T65A + Q327F + Y504T; T65A + E501V + Y504T; Q327F + E501V + Y504T; T65A + Q327F + Y504*; T65A + E501V + Y504*; Q327F + E501V + Y504*; T65A + Q327F + E501V + Y504T; T65A + Q327F + E501V + Y504*; E501V + Y504T; T65A + K161S; T65A + Q405T; T65A + Q327W; T65A + Q327F; T65A + Q327Y; P11F + T65A + Q327F; R1K + D3W + K5Q + G7V + N8S + T10K + P11S + T65A + Q327F; P2N + P4S + P11F + T65A + Q327F; P11F + D26C + K33C + T65A + Q327F; P2N + P4S + P11F + T65A + Q327W + E501V + Y504T; R1E + D3N + P4G + G6R + G7A + N8A + T10D+ P11D + T65A + Q327F; P11F + T65A + Q327W; P2N + P4S + P11F + T65A + Q327F + E501V + Y504T; P11F + T65A + Q327W + E501V + Y504T; T65A + Q327F + E501V + Y504T; T65A + S105P + Q327W; T65A + S105P + Q327F; T65A + Q327W + S364P; T65A + Q327F + S364P; T65A + S103N + Q327F; P2N + P4S + P11F + K34Y + T65A + Q327F; P2N + P4S + P11F + T65A + Q327F + D445N + V447S; P2N + P4S + P11F + T65A + I172V + Q327F; P2N + P4S + P11F + T65A + Q327F + N502*; P2N + P4S + P11F + T65A + Q327F + N502T + P563S + K571E; P2N + P4S + P11F + R31S + K33V + T65A + Q327F + N564D + K571S; P2N + P4S + P11F + T65A + Q327F + S377T; P2N + P4S + P11F + T65A + V325T+ Q327W; P2N + P4S + P11F + T65A + Q327F + D445N + V447S + E501V + Y504T; P2N + P4S + P11F + T65A + I172V + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + S377T + E501V + Y504T; P2N + P4S + P11F + D26N + K34Y + T65A + Q327F; P2N + P4S + P11F + T65A + Q327F + I375A + E501V + Y504T; P2N + P4S + P11F + T65A + K218A + K221D + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + S103N + Q327F + E501V + Y504T; P2N + P4S + T10D + T65A + Q327F + E501V + Y504T; P2N + P4S + F12Y + T65A + Q327F + E501V + Y504T; K5A + P11F + T65A + Q327F + E501V + Y504T; P2N + P4S + T10E + E18N + T65A + Q327F + E501V + Y504T; P2N + T10E + E18N + T65A + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + E501V + Y504T + T568N; P2N + P4S +

CA 03158982 2022-5-19

P11F + T65A + Q327F + E501V + Y504T + K524T + G526A; P2N + P4S + P11F + K34Y + T65A + Q327F + D445N + V447S + E501V + Y504T; P2N + P4S + P11F + R31S + K33V + T65A + Q327F + D445N + V447S + E501V + Y504T; P2N + P4S + P11F + D26N + K34Y + T65A + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + F80* + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + K112S + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + E501V + Y504T + T516P + K524T + G526A; P2N + P4S + P11F + T65A + Q327F + E501V + N502T + Y504*; P2N + P4S + P11F + T65A + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + S103N + Q327F + E501V + Y504T; K5A + P11F + T65A + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + E501V + Y504T + T516P + K524T + G526A; P2N + P4S + P11F + T65A + V79A + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + V79G + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + V79I + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + V79L + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + V79S + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + L72V + Q327F + E501V + Y504T; S255N + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + E74N + V79K + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + G220N + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + Y245N + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + Q253N + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + D279N + Q327F + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + S359N + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + D370N + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + V460S + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + V460T + P468T + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + T463N + E501V + Y504T; P2N + P4S + P11F + T65A + Q327F + S465N + E501V + Y504T; and P2N + P4S + P11F + T65A + Q327F + T477N + E501V + Y504T.

In one embodiment, the *Penicillium oxalicum* glucoamylase variant has a K79V substitution (using SEQ ID NO: 2 of WO 2011/127802 for numbering), corresponding to the PE001 variant, and further comprises one of the following substitutions or combinations of substitutions:

P11F + T65A + Q327F;

P2N + P4S + P11F + T65A + Q327F;

P11F + D26C + K33C + T65A + Q327F;

P2N + P4S + P11F + T65A + Q327W + E501V + Y504T;

P2N + P4S + P11F + T65A + Q327F + E501V + Y504T; and

P11F + T65A + Q327W + E501V + Y504T.

Additional glucoamylases contemplated for use with the present invention can be found in WO2011/153516 (the content of which is incorporated herein).

CA 03158982 2022-5-19

Additional polynucleotides encoding suitable glucoamylases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org).

The glucoamylase coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding glucoamylases from strains of different genera or species, as described *supra*.

The polynucleotides encoding glucoamylases may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc,) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding glucoamylases are described *supra*.

In one embodiment, the glucoamylase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any one of the glucoamylases described or referenced herein (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250). In another embodiment, the glucoamylase has a mature polypeptide sequence that is a fragment of the any one of the glucoamylases described or referenced herein (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250). In one embodiment, the number of amino acid residues in the fragment is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of amino acid residues in referenced full length glucoamylase (e.g. any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250). In other embodiments, the glucoamylase may comprise the catalytic domain of any glucoamylase described or referenced herein (e.g., the catalytic domain of any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250).

The glucoamylase may be a variant of any one of the glucoamylases described *supra* (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250). In one embodiment, the glucoamylase has a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of the glucoamylases described *supra* (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250).

Examples of suitable amino acid changes, such as conservative substitutions that do not significantly affect the folding and/or activity of the glucoamylase, are described herein.

In one embodiment, the glucoamylase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino

CA 03158982 2022-5-19

acid from the amino acid sequence of any one of the glucoamylases described *supra* (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250). In one embodiment, the glucoamylase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) of amino acid sequence of any one of the glucoamylases described *supra* (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250). In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the glucoamylase has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the glucoamylase activity of any glucoamylase described or referenced herein (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250) under the same conditions.

In one embodiment, the glucoamylase coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any glucoamylase described or referenced herein (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250). In one embodiment, the glucoamylase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any glucoamylase described or referenced herein (e.g., any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250).

In one embodiment, the glucoamylase comprises the coding sequence of any glucoamylase described or referenced herein (any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250). In one embodiment, the glucoamylase comprises a coding sequence that is a subsequence of the coding sequence from any glucoamylase described or referenced herein, wherein the subsequence encodes a polypeptide having glucoamylase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The referenced glucoamylase coding sequence of any related aspect or embodiment described herein can be the native coding sequence or a degenerate sequence, such as a codon-optimized coding sequence designed for use in a particular host cell (e.g., optimized for expression in *Saccharomyces cerevisiae*).

CA 03158982 2022-5-19

The glucoamylase can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

Alpha-Amylases

The host cells and fermenting organisms may express a heterologous alpha-amylase. The alpha-amylase may be any alpha-amylase that is suitable for the host cells and/or the methods described herein, such as a naturally occurring alpha-amylase (e.g., a native alpha-amylase from another species or an endogenous alpha-amylase expressed from a modified expression vector) or a variant thereof that retains alpha-amylase activity. Any alpha-amylase contemplated for expression by a host cell or fermenting organism described below is also contemplated for embodiments of the invention involving exogenous addition of an alpha-amylase.

In some embodiments, the host cell or fermenting organism comprises a heterologous polynucleotide encoding an alpha-amylase, for example, as described in WO2017/087330 or PCT/US2019/042870, the content of which is hereby incorporated by reference. Any alpha-amylase described or referenced herein is contemplated for expression in the host cell or fermenting organism.

In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding an alpha-amylase has an increased level of alpha-amylase activity compared to the host cells without the heterologous polynucleotide encoding the alpha-amylase, when cultivated under the same conditions. In some embodiments, the host cell or fermenting organism has an increased level of alpha-amylase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the alpha-amylase, when cultivated under the same conditions (e.g., as described in Example 2).

Exemplary alpha-amylases that can be used with the host cells and/or the methods described herein include bacterial, yeast, or filamentous fungal alpha-amylases, e.g., derived from any of the microorganisms described or referenced herein.

The term "bacterial alpha-amylase" means any bacterial alpha-amylase classified under EC 3.2.1.1. A bacterial alpha-amylase used herein may, e.g., be derived from a strain of the genus *Bacillus*, which is sometimes also referred to as the genus *Geobacillus*. In one embodiment, the *Bacillus* alpha-amylase is derived from a strain of *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, *Bacillus stearothermophilus*, or *Bacillus subtilis*, but may also be derived from other *Bacillus* sp.

CA 03158982 2022-5-19

Specific examples of bacterial alpha-amylases include the *Bacillus stearothermophilus* alpha-amylase (BSG) of SEQ ID NO: 3 in WO 99/19467, the *Bacillus amyloliquefaciens* alpha-amylase (BAN) of SEQ ID NO: 5 in WO 99/19467, and the *Bacillus licheniformis* alpha-amylase (BLA) of SEQ ID NO: 4 in WO 99/19467 (all sequences are hereby incorporated by reference). In one embodiment, the alpha-amylase may be an enzyme having a mature polypeptide sequence with a degree of identity of at least 60%, e.g., at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% to any of the sequences shown in SEQ ID NOs: 3, 4 or 5, in WO 99/19467.

In one embodiment, the alpha-amylase is derived from *Bacillus stearothermophilus*. The *Bacillus stearothermophilus* alpha-amylase may be a mature wild-type or a mature variant thereof. The mature *Bacillus stearothermophilus* alpha-amylases may naturally be truncated during recombinant production. For instance, the *Bacillus stearothermophilus* alpha-amylase may be a truncated at the C-terminal, so that it is from 480-495 amino acids long, such as about 491 amino acids long, e.g., so that it lacks a functional starch binding domain (compared to SEQ ID NO: 3 in WO 99/19467).

The *Bacillus* alpha-amylase may also be a variant and/or hybrid. Examples of such a variant can be found in any of WO 96/23873, WO 96/23874, WO 97/41213, WO 99/19467, WO 00/60059, and WO 02/10355 (each hereby incorporated by reference). Specific alpha-amylase variants are disclosed in U.S. Patent Nos. 6,093,562, 6,187,576, 6,297,038, and 7,713,723 (hereby incorporated by reference) and include *Bacillus stearothermophilus* alpha-amylase (often referred to as BSG alpha-amylase) variants having a deletion of one or two amino acids at positions R179, G180, I181 and/or G182, preferably a double deletion disclosed in WO 96/23873 – see, e.g., page 20, lines 1-10 (hereby incorporated by reference), such as corresponding to deletion of positions I181 and G182 compared to the amino acid sequence of *Bacillus stearothermophilus* alpha-amylase set forth in SEQ ID NO: 3 disclosed in WO 99/19467 or the deletion of amino acids R179 and G180 using SEQ ID NO: 3 in WO 99/19467 for numbering (which reference is hereby incorporated by reference). In some embodiments, the *Bacillus* alpha-amylases, such as *Bacillus stearothermophilus* alpha-amylases, have a double deletion corresponding to a deletion of positions 181 and 182 and further optionally comprise a N193F substitution (also denoted I181* + G182* + N193F) compared to the wild-type BSG alpha-amylase amino acid sequence set forth in SEQ ID NO: 3 disclosed in WO 99/19467. The bacterial alpha-amylase may also have a substitution in a position corresponding to S239 in the *Bacillus licheniformis* alpha-amylase shown in SEQ ID NO: 4 in WO 99/19467, or a S242 and/or E188P variant of the *Bacillus stearothermophilus* alpha-amylase of SEQ ID NO: 3 in WO 99/19467.

CA 03158982 2022-5-19

In one embodiment, the variant is a S242A, E or Q variant, e.g., a S242Q variant, of the *Bacillus stearothermophilus* alpha-amylase.

In one embodiment, the variant is a position E188 variant, e.g., E188P variant of the *Bacillus stearothermophilus* alpha-amylase.

The bacterial alpha-amylase may, in one embodiment, be a truncated *Bacillus* alpha-amylase. In one embodiment, the truncation is so that, e.g., the *Bacillus stearothermophilus* alpha-amylase shown in SEQ ID NO: 3 in WO 99/19467, is about 491 amino acids long, such as from 480 to 495 amino acids long, or so it lacks a functional starch bind domain.

The bacterial alpha-amylase may also be a hybrid bacterial alpha-amylase, e.g., an alpha-amylase comprising 445 C-terminal amino acid residues of the *Bacillus licheniformis* alpha-amylase (shown in SEQ ID NO: 4 of WO 99/19467) and the 37 N-terminal amino acid residues of the alpha-amylase derived from *Bacillus amyloliquefaciens* (shown in SEQ ID NO: 5 of WO 99/19467). In one embodiment, this hybrid has one or more, especially all, of the following substitutions: G48A+T49I+G107A+H156Y+A181T+N190F+I201F+A209V+Q264S (using the *Bacillus licheniformis* numbering in SEQ ID NO: 4 of WO 99/19467). In some embodiments, the variants have one or more of the following mutations (or corresponding mutations in other *Bacillus* alpha-amylases): H154Y, A181T, N190F, A209V and Q264S and/or the deletion of two residues between positions 176 and 179, e.g., deletion of E178 and G179 (using SEQ ID NO: 5 of WO 99/19467 for position numbering).

In one embodiment, the bacterial alpha-amylase is the mature part of the chimeric alpha-amylase disclosed in Richardson et al. (2002), The Journal of Biological Chemistry, Vol. 277, No 29, Issue 19 July, pp. 267501-26507, referred to as BD5088 or a variant thereof. This alpha-amylase is the same as the one shown in SEQ ID NO: 2 in WO 2007/134207. The mature enzyme sequence starts after the initial "Met" amino acid in position 1.

The alpha-amylase may be a thermostable alpha-amylase, such as a thermostable bacterial alpha-amylase, e.g., from *Bacillus stearothermophilus*. In one embodiment, the alpha-amylase used in a process described herein has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$ of at least 10 determined as described in Example 1 of WO2018/098381.

In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, of at least 15. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, of as at least 20. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, of as at least 25. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, of as at least

CA 03158982 2022-5-19

30. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, of as at least 40.

In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, of at least 50. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, of at least 60. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, between 10-70. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, between 15-70. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, between 20-70. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, between 25-70. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, between 30-70. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, between 40-70. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, between 50-70. In one embodiment, the thermostable alpha-amylase has a T½ (min) at pH 4.5, 85°C, 0.12 mM $CaCl_2$, between 60-70.

In one embodiment, the alpha-amylase is a bacterial alpha-amylase, e.g., derived from the genus *Bacillus*, such as a strain of *Bacillus stearothermophilus,* e.g., the *Bacillus stearothermophilus* as disclosed in WO 99/019467 as SEQ ID NO: 3 with one or two amino acids deleted at positions R179, G180, I181 and/or G182, in particular with R179 and G180 deleted, or with I181 and G182 deleted, with mutations in below list of mutations.

In some embodiment, the *Bacillus stearothermophilus* alpha-amylases have double deletion I181 + G182, and optional substitution N193F, further comprising one of the following substitutions or combinations of substitutions:

V59A+Q89R+G112D+E129V+K177L+R179E+K220P+N224L+Q254S;
V59A+Q89R+E129V+K177L+R179E+H208Y+K220P+N224L+Q254S;
V59A+Q89R+E129V+K177L+R179E+K220P+N224L+Q254S+D269E+D281N;
V59A+Q89R+E129V+K177L+R179E+K220P+N224L+Q254S+I270L;
V59A+Q89R+E129V+K177L+R179E+K220P+N224L+Q254S+H274K;
V59A+Q89R+E129V+K177L+R179E+K220P+N224L+Q254S+Y276F;
V59A+E129V+R157Y+K177L+R179E+K220P+N224L+S242Q+Q254S;
V59A+E129V+K177L+R179E+H208Y+K220P+N224L+S242Q+Q254S;
V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S;
V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+H274K;
V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+Y276F;

V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+D281N;
V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+M284T;
V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+G416V;
V59A+E129V+K177L+R179E+K220P+N224L+Q254S;
V59A+E129V+K177L+R179E+K220P+N224L+Q254S+M284T;
A91L+M96I+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S;
E129V+K177L+R179E;
E129V+K177L+R179E+K220P+N224L+S242Q+Q254S;
E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+Y276F+L427M;
E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+M284T;
E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+N376*+I377*;
E129V+K177L+R179E+K220P+N224L+Q254S;
E129V+K177L+R179E+K220P+N224L+Q254S+M284T;
E129V+K177L+R179E+S242Q;
E129V+K177L+R179V+K220P+N224L+S242Q+Q254S;
K220P+N224L+S242Q+Q254S;
M284V;
V59A+Q89R+ E129V+ K177L+ R179E+ Q254S+ M284V; and
V59A+E129V+K177L+R179E+Q254S+ M284V;

In one embodiment, the alpha-amylase is selected from the group of *Bacillus stearothermophilus* alpha-amylase variants with double deletion I181*+G182*, and optionally substitution N193F, and further one of the following substitutions or combinations of substitutions:

E129V+K177L+R179E;
V59A+Q89R+E129V+K177L+R179E+H208Y+K220P+N224L+Q254S;
V59A+Q89R+ E129V+ K177L+ R179E+ Q254S+ M284V;
V59A+E129V+K177L+R179E+Q254S+ M284V; and
E129V+K177L+R179E+K220P+N224L+S242Q+Q254S (using SEQ ID NO: 1 herein for numbering).

It should be understood that when referring to *Bacillus stearothermophilus* alpha-amylase and variants thereof they are normally produced in truncated form. In particular, the truncation may be so that the *Bacillus stearothermophilus* alpha-amylase shown in SEQ ID NO: 3 in WO 99/19467, or variants thereof, are truncated in the C-terminal and are typically from 480-495 amino acids long, such as about 491 amino acids long, e.g., so that it lacks a functional starch binding domain.

CA 03158982 2022-5-19

In one embodiment, the alpha-amylase variant may be an enzyme having a mature polypeptide sequence with a degree of identity of at least 60%, e.g., at least 70%, at least 80%, at least 90%, at least 95%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99%, but less than 100% to the sequence shown in SEQ ID NO: 3 in WO 99/19467.

In one embodiment, the bacterial alpha-amylase, e.g., *Bacillus* alpha-amylase, such as especially *Bacillus stearothermophilus* alpha-amylase, or variant thereof, is dosed to liquefaction in a concentration between 0.01-10 KNU-A/g DS, e.g., between 0.02 and 5 KNU-A/g DS, such as 0.03 and 3 KNU-A, preferably 0.04 and 2 KNU-A/g DS, such as especially 0.01 and 2 KNU-A/g DS. In one embodiment, the bacterial alpha-amylase, e.g., *Bacillus* alpha-amylase, such as especially *Bacillus stearothermophilus* alpha-amylases, or variant thereof, is dosed to liquefaction in a concentration of between 0.0001-1 mg EP (Enzyme Protein)/g DS, e.g., 0.0005-0.5 mg EP/g DS, such as 0.001-0.1 mg EP/g DS.

In one embodiment, the bacterial alpha-amylase is derived from the *Bacillus subtilis* alpha-amylase of SEQ ID NO: 76, the *Bacillus subtilis* alpha-amylase of SEQ ID NO: 82, the *Bacillus subtilis* alpha-amylase of SEQ ID NO: 83, the *Bacillus subtilis* alpha-amylase of SEQ ID NO: 84, or the *Bacillus licheniformis* alpha-amylase of SEQ ID NO: 85, the *Clostridium phytofermentans* alpha-amylase of SEQ ID NO: 89, the *Clostridium phytofermentans* alpha-amylase of SEQ ID NO: 90, the *Clostridium phytofermentans* alpha-amylase of SEQ ID NO: 91, the *Clostridium phytofermentans* alpha-amylase of SEQ ID NO: 92, the *Clostridium phytofermentans* alpha-amylase of SEQ ID NO: 93, the *Clostridium phytofermentans* alpha-amylase of SEQ ID NO: 94, the *Clostridium thermocellum* alpha-amylase of SEQ ID NO: 95, the *Thermobifida fusca* alpha-amylase of SEQ ID NO: 96, the *Thermobifida fusca* alpha-amylase of SEQ ID NO: 97, the *Anaerocellum thermophilum* of SEQ ID NO: 98, the *Anaerocellum thermophilum* of SEQ ID NO: 99, the *Anaerocellum thermophilum* of SEQ ID NO: 100, the *Streptomyces avermitilis* of SEQ ID NO: 101, or the *Streptomyces avermitilis* of SEQ ID NO: 88.

In one embodiment, the alpha-amylase is derived from *Bacillus amyloliquefaciens*, such as the *Bacillus amyloliquefaciens* alpha-amylase of SEQ ID NO: 231 (e.g., as described in WO2018/002360, or variants thereof as described in WO2017/037614).

In one embodiment, the alpha-amylase is derived from a yeast alpha-amylase, such as the *Saccharomycopsis fibuligera* alpha-amylase of SEQ ID NO: 77, the *Debaryomyces occidentalis* alpha-amylase of SEQ ID NO: 78, the *Debaryomyces occidentalis* alpha-amylase of SEQ ID NO: 79, the *Lipomyces kononenkoae* alpha-amylase of SEQ ID NO: 80, the *Lipomyces kononenkoae* alpha-amylase of SEQ ID NO: 81.

CA 03158982 2022-5-19

In one embodiment, the alpha-amylase is derived from a filamentous fungal alpha-amylase, such as the *Aspergillus niger* alpha-amylase of SEQ ID NO: 86, or the *Aspergillus niger* alpha-amylase of SEQ ID NO: 87.

Additional alpha-amylases that may be expressed with the host cells and fermenting organisms and used with the methods described herein are described in the examples, and include, but are not limited to alpha-amylases shown in Table 3 (or derivatives thereof).

Table 3.

Donor Organism (catalytic domain)	SEQ ID NO: (mature polypeptide)
Rhizomucor pusillus	121
Bacillus licheniformis	122
Aspergillus niger	123
Aspergillus tamarii	124
Acidomyces richmondensis	125
Aspergillus bombycis	126
Alternaria sp	127
Rhizopus microsporus	128
Syncephalastrum racemosum	129
Rhizomucor pusillus	130
Dichotomocladium hesseltinei	131
Lichtheimia ramosa	132
Penicillium aethiopicum	133
Subulispora sp	134
Trichoderma paraviridescens	135
Byssoascus striatosporus	136
Aspergillus brasiliensis	137
Penicillium subspinulosum	138
Penicillium antarcticum	139
Penicillium coprophilum	140
Penicillium olsonii	141
Penicillium vasconiae	142
Penicillium sp	143
Heterocephalum aurantiacum	144
Neosartorya massa	145
Penicillium janthinellum	146
Aspergillus brasiliensis	147
Aspergillus westerdijkiae	148
Hamigera avellanea	149
Hamigera avellanea	150

CA 03158982 2022-5-19

Meripilus giganteus	151
Cerrena unicolor	152
Physalacria cryptomeriae	153
Lenzites betulinus	154
Trametes ljubarskyi	155
Bacillus subtilis	156
Bacillus subtilis subsp. subtilis	157
Schwanniomyces occidentalis	158
Rhizomucor pusillus	159
Aspergillus niger	160
Bacillus stearothermophilus	161
Bacillus halmapalus	162
Aspergillus oryzae	163
Bacillus amyloliquefaciens	164
Rhizomucor pusillus	165
Kionochaeta ivoriensis	166
Aspergillus niger	167
Aspergillus oryzae	168
Penicillium canescens	169
Acidomyces acidothermus	170
Kinochaeta ivoriensis	171
Aspergillus terreus	172
Thamnidium elegans	173
Meripilus giganteus	174
Bacillus amyloliquefaciens	231
Thermococcus gammatolerans	251
Thermococcus thioreducens	252
Thermococcus eurythermalis	253
Thermococcus hydrothermalis	254
Pyrococcus furiosus	255
Bacillus amyloliquefaciens	256

Additional alpha-amylases contemplated for use with the present invention can be found in WO2011/153516, WO2017/087330 and PCT/US2019/042870 (the content of which is incorporated herein).

Additional polynucleotides encoding suitable alpha-amylases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org).

CA 03158982 2022-5-19

The alpha-amylase coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding trehalases from strains of different genera or species, as described *supra*.

The polynucleotides encoding alpha-amylases may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding alpha-amylases are described *supra*.

In one embodiment, the alpha-amylase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any one of the alpha-amylases described or referenced herein (e.g., any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256). In another embodiment, the alpha-amylase has a mature polypeptide sequence that is a fragment of the any one of the alpha-amylases described or referenced herein (e.g., any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256). In one embodiment, the number of amino acid residues in the fragment is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of amino acid residues in referenced full length alpha-amylase (e.g. any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256). In other embodiments, the alpha-amylase may comprise the catalytic domain of any alpha-amylase described or referenced herein (e.g., the catalytic domain of any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256).

The alpha-amylase may be a variant of any one of the alpha-amylases described *supra* (e.g., any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256). In one embodiment, the alpha-amylase has a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of the alpha-amylases described *supra* (e.g., any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256).

Examples of suitable amino acid changes, such as conservative substitutions that do not significantly affect the folding and/or activity of the alpha-amylase, are described herein.

In one embodiment, the alpha-amylase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from the amino acid sequence of any one of the alpha-amylases described *supra* (e.g., any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256). In one embodiment, the alpha-amylase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) of amino acid sequence of any one of the alpha-amylases described *supra* (e.g., any one of SEQ

CA 03158982 2022-5-19

ID NOs: 76-101, 121-174, 231 and 251-256). In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the alpha-amylase has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the alpha-amylase activity of any alpha-amylase described or referenced herein (e.g., any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256) under the same conditions.

In one embodiment, the alpha-amylase coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any alpha-amylase described or referenced herein (e.g., any one of SEQ ID NOs: 76-101, 121-174 and 231). In one embodiment, the alpha-amylase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any alpha-amylase described or referenced herein (e.g., any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256).

In one embodiment, the alpha-amylase comprises the coding sequence of any alpha-amylase described or referenced herein (any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256). In one embodiment, the alpha-amylase comprises a coding sequence that is a subsequence of the coding sequence from any alpha-amylase described or referenced herein, wherein the subsequence encodes a polypeptide having alpha-amylase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The referenced alpha-amylase coding sequence of any related aspect or embodiment described herein can be the native coding sequence or a degenerate sequence, such as a codon-optimized coding sequence designed for use in a particular host cell (e.g., optimized for expression in *Saccharomyces cerevisiae*).

The alpha-amylase can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

CA 03158982 2022-5-19

Phospholipases

The host cells and fermenting organisms may express a heterologous phospholipase. The phospholipase may be any phospholipase that is suitable for the host cells, fermenting organism, and/or the methods described herein, such as a naturally occurring phospholipase (e.g., a native phospholipase from another species or an endogenous phospholipase expressed from a modified expression vector) or a variant thereof that retains phospholipase activity. Any phospholipase contemplated for expression by a host cell or fermenting organism described below is also contemplated for embodiments of the invention involving exogenous addition of a phospholipase (e.g., added before, during or after liquefaction and/or saccharification).

In some embodiments, the host cell or fermenting organism comprises a heterologous polynucleotide encoding a phospholipase, for example, as described in WO2018/075430, the content of which is hereby incorporated by reference. In some embodiments, the phospholipase is classified as a phospholipase A. In other embodiments, the phospholipase is classified as a phospholipase C. Any phospholipase described or referenced herein is contemplated for expression in the host cell or fermenting organism.

In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding a phospholipase has an increased level of phospholipase activity compared to the host cells without the heterologous polynucleotide encoding the phospholipase, when cultivated under the same conditions. In some embodiments, the host cell or fermenting organism has an increased level of phospholipase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the phospholipase, when cultivated under the same conditions.

Exemplary phospholipases that can be used with the host cells and/or the methods described herein include bacterial, yeast, or filamentous fungal phospholipases, e.g., derived from any of the microorganisms described or referenced herein.

Additional phospholipases that may be expressed with the host cells and fermenting organisms, and used with the methods described herein, and include, but are not limited to phospholipases shown in Table 4 (or derivatives thereof).

Table 4.

Donor Organism (catalytic domain)	SEQ ID NO: (mature polypeptide)
Thermomyces lanuginosus	235
Talaromyces leycettanus	236

Penicillium emersonii	237
Bacillus thuringiensis	238
Pseudomonas sp.	239
Kionochaeta sp.	240
Mariannaea pinicola	241
Fictibacillus macauensis	242

Additional phospholipases contemplated for use with the present invention can be found in WO2018/075430 (the content of which is incorporated herein).

Additional polynucleotides encoding suitable phospholipases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org).

The phospholipase coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding phospholipases from strains of different genera or species, as described *supra*.

The polynucleotides encoding phospholipases may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding phospholipases are described *supra*.

In one embodiment, the phospholipase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any one of the phospholipases described or referenced herein (e.g., any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242). In another embodiment, the phospholipase has a mature polypeptide sequence that is a fragment of the any one of the phospholipases described or referenced herein (e.g., any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242). In one embodiment, the number of amino acid residues in the fragment is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of amino acid residues in referenced full length phospholipase (e.g. any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242). In other embodiments, the phospholipase may comprise the catalytic domain of any phospholipase described or referenced herein (e.g., the catalytic domain of any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242).

The phospholipase may be a variant of any one of the phospholipases described *supra* (e.g., any one of SEQ ID NOs: SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242). In one embodiment, the phospholipase has a mature polypeptide sequence of at least 60%, e.g., at least

65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of the phospholipases described *supra* (e.g., any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242).

Examples of suitable amino acid changes, such as conservative substitutions that do not significantly affect the folding and/or activity of the phospholipase, are described herein.

In one embodiment, the phospholipase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from the amino acid sequence of any one of the phospholipases described *supra* (e.g., any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242). In one embodiment, the phospholipase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) of amino acid sequence of any one of the phospholipases described *supra* (e.g., any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242). In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the phospholipase has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the phospholipase activity of any phospholipase described or referenced herein (e.g., any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241, and 242) under the same conditions.

In one embodiment, the phospholipase coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any phospholipase described or referenced herein (e.g., a coding sequence for a phospholipase of SEQ ID NO: 235, 236, 237, 238, 239, 240, 241 or 242). In one embodiment, the phospholipase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any phospholipase described or referenced herein (e.g., a coding sequence for a phospholipase of SEQ ID NO: 235, 236, 237, 238, 239, 240, 241 or 242).

In one embodiment, the phospholipase comprises the coding sequence of any phospholipase described or referenced herein (e.g., a coding sequence for a phospholipase of SEQ ID NO: 235, 236, 237, 238, 239, 240, 241 or 242). In one embodiment, the phospholipase

CA 03158982 2022-5-19

comprises a coding sequence that is a subsequence of the coding sequence from any phospholipase described or referenced herein, wherein the subsequence encodes a polypeptide having phospholipase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The referenced phospholipase coding sequence of any related aspect or embodiment described herein can be the native coding sequence or a degenerate sequence, such as a codon-optimized coding sequence designed for use in a particular host cell (e.g., optimized for expression in *Saccharomyces cerevisiae*).

The phospholipase can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

Trehalases

The host cells and fermenting organisms may express a heterologous trehalase. The trehalase can be any trehalase that is suitable for the host cells, fermenting organisms and/or their methods of use described herein, such as a naturally occurring trehalase or a variant thereof that retains trehalase activity. Any trehalase contemplated for expression by a host cell or fermenting organism described below is also contemplated for embodiments of the invention involving exogenous addition of a trehalase (e.g., added before, during or after liquefaction and/or saccharification).

In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding a trehalase has an increased level of trehalase activity compared to the host cells without the heterologous polynucleotide encoding the trehalase, when cultivated under the same conditions. In some embodiments, the host cell or fermenting organism has an increased level of trehalase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the trehalase, when cultivated under the same conditions.

Trehalases that may be expressed with the host cells and fermenting organisms, and used with the methods described herein include, but are not limited to, trehalases shown in Table 5 (or derivatives thereof).

Table 5.

Donor Organism (catalytic domain)	SEQ ID NO: (mature polypeptide)

Chaetomium megalocarpum	175
Lecanicillium psalliotae	176
Doratomyces sp	177
Mucor moelleri	178
Phialophora cyclaminis	179
Thielavia arenaria	180
Thielavia antarctica	181
Chaetomium sp	182
Chaetomium nigricolor	183
Chaetomium jodhpurense	184
Chaetomium piluliferum	185
Myceliophthora hinnulea	186
Chloridium virescens	187
Gelasinospora cratophora	188
Acidobacteriaceae bacterium	189
Acidobacterium capsulatum	190
Acidovorax wautersii	191
Xanthomonas arborícola	192
Kosakonia sacchari	193
Enterobacter sp	194
Saitozyma flava	195
Phaeotremella skinneri	196
Trichoderma asperellum	197
Corynascus sepedonium	198
Myceliophthora thermophila	199
Trichoderma reesei	200
Chaetomium virescens	201
Rhodothermus marinus	202
Myceliophthora sepedonium	203
Moelleriella libera	204
Acremonium dichromosporum	205
Fusarium sambucinum	206
Phoma sp	207
Lentinus similis	208
Diaporthe nobilis	209
Solicoccozyma terricola	210
Dioszegia cryoxerica	211
Talaromyces funiculosus	212
Hamigera avellanea	213
Talaromyces ruber	214
Trichoderma lixii	215

Aspergillus cervinus	216
Rasamsonia brevistipitata	217
Acremonium curvulum	218
Talaromyces piceae	219
Penicillium sp	220
Talaromyces aurantiacus	221
Talaromyces pinophilus	222
Talaromyces leycettanus	223
Talaromyces variabilis	224
Aspergillus niger	225
Trichoderma reesei	226

Additional polynucleotides encoding suitable trehalases may be derived from microorganisms of any suitable genus, including those readily available within the UniProtKB database (www.uniprot.org).

The trehalase coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding trehalases from strains of different genera or species, as described *supra*.

The polynucleotides encoding trehalases may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding trehalases are described *supra*.

In one embodiment, the trehalase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any one of the trehalases described or referenced herein (e.g., any one of SEQ ID NOs: 175-226). In another embodiment, the trehalase has a mature polypeptide sequence that is a fragment of the any one of the trehalases described or referenced herein (e.g., any one of SEQ ID NOs: 175-226). In one embodiment, the number of amino acid residues in the fragment is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of amino acid residues in referenced full length trehalase (e.g. any one of SEQ ID NOs: 175-226). In other embodiments, the trehalase may comprise the catalytic domain of any trehalase described or referenced herein (e.g., the catalytic domain of any one of SEQ ID NOs: 175-226).

The trehalase may be a variant of any one of the trehalases described *supra* (e.g., any one of SEQ ID NOs: 175-226). In one embodiment, the trehalase has a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%,

CA 03158982 2022-5-19

or 100% sequence identity to any one of the trehalases described *supra* (e.g., any one of SEQ ID NOs: 175-226).

Examples of suitable amino acid changes, such as conservative substitutions that do not significantly affect the folding and/or activity of the trehalase, are described herein.

In one embodiment, the trehalase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from the amino acid sequence of any one of the trehalases described *supra* (e.g., any one of SEQ ID NOs: 175-226). In one embodiment, the trehalase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) of amino acid sequence of any one of the trehalases described *supra* (e.g., any one of SEQ ID NOs: 175-226). In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the trehalase has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the trehalase activity of any trehalase described or referenced herein (e.g., any one of SEQ ID NOs: 175-226) under the same conditions.

In one embodiment, the trehalase coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any trehalase described or referenced herein (e.g., any one of SEQ ID NOs: 175-226). In one embodiment, the trehalase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any trehalase described or referenced herein (e.g., any one of SEQ ID NOs: 175-226).

In one embodiment, the trehalase comprises the coding sequence of any trehalase described or referenced herein (any one of SEQ ID NOs: 175-226). In one embodiment, the trehalase comprises a coding sequence that is a subsequence of the coding sequence from any trehalase described or referenced herein, wherein the subsequence encodes a polypeptide having trehalase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

CA 03158982 2022-5-19

The referenced trehalase coding sequence of any related aspect or embodiment described herein can be the native coding sequence or a degenerate sequence, such as a codon-optimized coding sequence designed for use in a particular host cell (e.g., optimized for expression in *Saccharomyces cerevisiae*).

The trehalase can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

Proteases

The host cells and fermenting organisms may express a heterologous protease. The protease can be any protease that is suitable for the host cells and fermenting organisms and/or their methods of use described herein, such as a naturally occurring protease or a variant thereof that retains protease activity. Any protease contemplated for expression by a host cell or fermenting organism described below is also contemplated for embodiments of the invention involving exogenous addition of a protease (e.g., added before, during or after liquefaction and/or saccharification).

Proteases are classified on the basis of their catalytic mechanism into the following groups: Serine proteases (S), Cysteine proteases (C), Aspartic proteases (A), Metallo proteases (M), and Unknown, or as yet unclassified, proteases (U), see Handbook of Proteolytic Enzymes, A.J.Barrett, N.D.Rawlings, J.F.Woessner (eds), Academic Press (1998), in particular the general introduction part.

Protease activity can be measured using any suitable assay, in which a substrate is employed, that includes peptide bonds relevant for the specificity of the protease in question. Assay-pH and assay-temperature are likewise to be adapted to the protease in question. Examples of assay-pH-values are pH 6, 7, 8, 9, 10, or 11. Examples of assay-temperatures are 30, 35, 37, 40, 45, 50, 55, 60, 65, 70 or 80°C.

In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding a protease has an increased level of protease activity compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the protease, when cultivated under the same conditions. In some embodiments, the host cell or fermenting organism has an increased level of protease activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the protease, when cultivated under the same conditions.

CA 03158982 2022-5-19

Exemplary proteases that may be expressed with the host cells and fermenting organisms, and used with the methods described herein include, but are not limited to, proteases shown in Table 6 (or derivatives thereof).

Table 6.

Donor Organism (catalytic domain)	SEQ ID NO: (mature polypeptide)	Family
Aspergillus niger	9	A1
Trichoderma reesei	10	
Thermoascus aurantiacus	11	M35
Dichomitus squalens	12	S53
Nocardiopsis prasina	13	S1
Penicillium simplicissimum	14	S10
Aspergillus niger	15	
Meriphilus giganteus	16	S53
Lecanicillium sp. WMM742	17	S53
Talaromyces proteolyticus	18	S53
Penicillium ranomafanaense	19	A1A
Aspergillus oryzae	20	S53
Talaromyces liani	21	S10
Thermoascus thermophilus	22	S53
Pyrococcus furiosus	23	
Trichoderma reesei	24	
Rhizomucor miehei	25	
Lenzites betulinus	26	S53
Neolentinus lepideus	27	S53
Thermococcus sp.	28	S8
Thermococcus sp.	29	S8
Thermomyces lanuginosus	30	S53
Thermococcus thioreducens	31	S53
Polyporus arcularius	32	S53
Ganoderma lucidum	33	S53
Ganoderma lucidum	34	S53
Ganoderma lucidum	35	S53
Trametes sp. AH28-2	36	S53
Cinereomyces lindbladii	37	S53
Trametes versicolor O82DDP	38	S53
Paecilomyces hepiali	39	S53
Isaria tenuipes	40	S53
Aspergillus tamarii	41	S53
Aspergillus brasiliensis	42	S53
Aspergillus iizukae	43	S53
Penicillium sp-72364	44	S10
Aspergillus denticulatus	45	S10
Hamigera sp. t184-6	46	S10
Penicillium janthinellum	47	S10
Penicillium vasconiae	48	S10

CA 03158982 2022-5-19

Hamigera paravellanea	49	S10
Talaromyces variabilis	50	S10
Penicillium arenicola	51	S10
Nocardiopsis kunsanensis	52	S1
Streptomyces parvulus	53	S1
Saccharopolyspora endophytica	54	S1
luteus cell wall enrichments K	55	S1
Saccharothrix australiensis	56	S1
Nocardiopsis baichengensis	57	S1
Streptomyces sp. SM15	58	S1
Actinoalloteichus spitiensis	59	S1
Byssochlamys verrucosa	60	M35
Hamigera terricola	61	M35
Aspergillus tamarii	62	M35
Aspergillus niveus	63	M35
Penicillium sclerotiorum	64	A1
Penicillium bilaiae	65	A1
Penicillium antarcticum	66	A1
Penicillium sumatrense	67	A1
Trichoderma lixii	68	A1
Trichoderma brevicompactum	69	A1
Penicillium cinnamopurpureum	70	A1
Bacillus licheniformis	71	S8
Bacillus subtilis	72	S8
Trametes cf versicol	73	S53

Additional polynucleotides encoding suitable proteases may be derived from microorganisms of any suitable genus, including those readily available within the UniProtKB database (www.uniprot.org).

In one embodiment, the protease is derived from *Aspergillus*, such as the *Aspergillus niger* protease of SEQ ID NO: 9, the *Aspergillus tamarii* protease of SEQ ID NO: 41, or the *Aspergillus denticulatus* protease of SEQ ID NO: 45. In one embodiment, the protease is derived from *Dichomitus*, such as the *Dichomitus squalens* protease of SEQ ID NO: 12. In one embodiment, the protease is derived from *Penicillium*, such as the *Penicillium simplicissimum* protease of SEQ ID NO: 14, the *Penicillium antarcticum* protease of SEQ ID NO: 66, or the *Penicillium sumatrense* protease of SEQ ID NO: 67. In one embodiment, the protease is derived from *Meriphilus*, such as the *Meriphilus giganteus* protease of SEQ ID NO: 16. In one embodiment, the protease is derived from *Talaromyces*, such as the *Talaromyces liani* protease of SEQ ID NO: 21. In one embodiment, the protease is derived from *Thermoascus*, such as the *Thermoascus thermophilus*

CA 03158982 2022-5-19

protease of SEQ ID NO: 22. In one embodiment, the protease is derived from *Ganoderma*, such as the *Ganoderma lucidum* protease of SEQ ID NO: 33. In one embodiment, the protease is derived from *Hamigera*, such as the *Hamigera terricola* protease of SEQ ID NO: 61. In one embodiment, the protease is derived from *Trichoderma*, such as the *Trichoderma brevicompactum* protease of SEQ ID NO: 69.

The protease coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding proteases from strains of different genera or species, as described *supra*.

The polynucleotides encoding proteases may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding proteases are described *supra*.

In one embodiment, the protease has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any one of SEQ ID NOs: 9-73 (e.g., any one of SEQ ID NOs: 9, 14, 16, 21, 22, 33, 41, 45, 61, 62, 66, 67, and 69; such as any one of SEQ NOs: 9, 14, 16, and 69). In another embodiment, the protease has a mature polypeptide sequence that is a fragment of the protease of any one of SEQ ID NOs: 9-73 (e.g., wherein the fragment has protease activity). In one embodiment, the number of amino acid residues in the fragment is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of amino acid residues in referenced full length protease (e.g. any one of SEQ ID NOs: 9-73). In other embodiments, the protease may comprise the catalytic domain of any protease described or referenced herein (e.g., the catalytic domain of any one of SEQ ID NOs: 9-73).

The protease may be a variant of any one of the proteases described *supra* (e.g., any one of SEQ ID NOs: 9-73. In one embodiment, the protease has a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of the proteases described *supra* (e.g., any one of SEQ ID NOs: 9-73).

Examples of suitable amino acid changes, such as conservative substitutions that do not significantly affect the folding and/or activity of the protease, are described herein.

In one embodiment, the protease has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino

CA 03158982 2022-5-19

acid from the amino acid sequence of any one of the proteases described *supra* (e.g., any one of SEQ ID NOs: 9-73). In one embodiment, the protease has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) of amino acid sequence of any one of the proteases described *supra* (e.g., any one of SEQ ID NOs: 9-73). In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In one embodiment, the protease coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any protease described or referenced herein (e.g., any one of SEQ ID NOs: 9-73). In one embodiment, the protease coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any protease described or referenced herein (e.g., any one of SEQ ID NOs: 9-73).

In one embodiment, the protease comprises the coding sequence of any protease described or referenced herein (any one of SEQ ID NOs: 9-73). In one embodiment, the protease comprises a coding sequence that is a subsequence of the coding sequence from any protease described or referenced herein, wherein the subsequence encodes a polypeptide having protease activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The referenced protease coding sequence of any related aspect or embodiment described herein can be the native coding sequence or a degenerate sequence, such as a codon-optimized coding sequence designed for use in a particular host cell (e.g., optimized for expression in *Saccharomyces cerevisiae*).

The protease can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

In one embodiment, the protease used according to a process described herein is a Serine proteases. In one particular embodiment, the protease is a serine protease belonging to the family 53, e.g., an endo-protease, such as S53 protease from *Meriphilus giganteus, Dichomitus squalens Trametes versicolor*, *Polyporus arcularius, Lenzites betulinus, Ganoderma lucidum, Neolentinus lepideus,* or *Bacillus* sp. 19138, in a process for producing ethanol from a starch-containing material, the ethanol yield was improved, when the S53 protease was present/or added during saccharification and/or fermentation of either gelatinized or un-gelatinized starch. In one

embodiment, the proteases is selected from: (a) proteases belonging to the EC 3.4.21 enzyme group; and/or (b) proteases belonging to the EC 3.4.14 enzyme group; and/or (c) Serine proteases of the peptidase family S53 that comprises two different types of peptidases: tripeptidyl aminopeptidases (exo-type) and endo-peptidases; as described in 1993, *Biochem. J.* 290:205-218 and in MEROPS protease database, release, 9.4 (31 January 2011) (www.merops.ac.uk). The database is described in Rawlings, N.D., Barrett, A.J. and Bateman, A., 2010, "MEROPS: the peptidase database", *Nucl. Acids Res.* 38: D227-D233.

For determining whether a given protease is a Serine protease, and a family S53 protease, reference is made to the above Handbook and the principles indicated therein. Such determination can be carried out for all types of proteases, be it naturally occurring or wild-type proteases; or genetically engineered or synthetic proteases.

Peptidase family S53 contains acid-acting endopeptidases and tripeptidyl-peptidases. The residues of the catalytic triad are Glu, Asp, Ser, and there is an additional acidic residue, Asp, in the oxyanion hole. The order of the residues is Glu, Asp, Asp, Ser. The Ser residue is the nucleophile equivalent to Ser in the Asp, His, Ser triad of subtilisin, and the Glu of the triad is a substitute for the general base, His, in subtilisin.

The peptidases of the S53 family tend to be most active at acidic pH (unlike the homologous subtilisins), and this can be attributed to the functional importance of carboxylic residues, notably Asp in the oxyanion hole. The amino acid sequences are not closely similar to those in family S8 (i.e. serine endopeptidase subtilisins and homologues), and this, taken together with the quite different active site residues and the resulting lower pH for maximal activity, provides for a substantial difference to that family. Protein folding of the peptidase unit for members of this family resembles that of subtilisin, having the clan type SB.

In one embodiment, the protease used according to a process described herein is a Cysteine proteases.

In one embodiment, the protease used according to a process described herein is a Aspartic proteases. Aspartic acid proteases are described in, for example, Hand-book of Proteolytic En-zymes, Edited by A.J. Barrett, N.D. Rawlings and J.F. Woessner, Aca-demic Press, San Diego, 1998, Chapter 270). Suitable examples of aspartic acid protease include, e.g., those disclosed in R.M. Berka et al. Gene, 96, 313 (1990)); (R.M. Berka et al. Gene, 125, 195-198 (1993)); and Gomi et al. Biosci. Biotech. Biochem. 57, 1095-1100 (1993), which are hereby incorporated by reference.

The protease also may be a metalloprotease, which is defined as a protease selected from the group consisting of:

(a) proteases belonging to EC 3.4.24 (metalloendopeptidases); preferably EC 3.4.24.39 (acid metallo proteinases);

(b) metalloproteases belonging to the M group of the above Handbook;

(c) metalloproteases not yet assigned to clans (designation: Clan MX), or belonging to either one of clans MA, MB, MC, MD, ME, MF, MG, MH (as defined at pp. 989-991 of the above Handbook);

(d) other families of metalloproteases (as defined at pp. 1448-1452 of the above Handbook);

(e) metalloproteases with a HEXXH motif;

(f) metalloproteases with an HEFTH motif;

(g) metalloproteases belonging to either one of families M3, M26, M27, M32, M34, M35, M36, M41, M43, or M47 (as defined at pp. 1448-1452 of the above Handbook);

(h) metalloproteases belonging to the M28E family; and

(i) metalloproteases belonging to family M35 (as defined at pp. 1492-1495 of the above Handbook).

In other particular embodiments, metalloproteases are hydrolases in which the nucleophilic attack on a peptide bond is mediated by a water molecule, which is activated by a divalent metal cation. Examples of divalent cations are zinc, cobalt or manganese. The metal ion may be held in place by amino acid ligands. The number of ligands may be five, four, three, two, one or zero. In a particular embodiment the number is two or three, preferably three.

There are no limitations on the origin of the metalloprotease used in a process of the invention. In an embodiment the metalloprotease is classified as EC 3.4.24, preferably EC 3.4.24.39. In one embodiment, the metalloprotease is an acid-stable metalloprotease, e.g., a fungal acid-stable metalloprotease, such as a metalloprotease derived from a strain of the genus *Thermoascus*, preferably a strain of *Thermoascus aurantiacus*, especially *Thermoascus aurantiacus* CGMCC No. 0670 (classified as EC 3.4.24.39). In another embodiment, the metalloprotease is derived from a strain of the genus *Aspergillus*, preferably a strain of *Aspergillus oryzae*.

In one embodiment the metalloprotease has a degree of sequence identity to amino acids -178 to 177, -159 to 177, or preferably amino acids 1 to 177 (the mature polypeptide) of SEQ ID NO: 1 of WO 2010/008841 (a *Thermoascus aurantiacus* metalloprotease) of at least 80%, at least 82%, at least 85%, at least 90%, at least 95%, or at least 97%; and which have metalloprotease activity. In particular embodiments, the metalloprotease consists of an amino acid sequence with a degree of identity to SEQ ID NO: 1 as mentioned above.

The *Thermoascus aurantiacus* metalloprotease is a preferred example of a metalloprotease suitable for use in a process of the invention. Another metalloprotease is derived from *Aspergillus oryzae* and comprises the sequence of SEQ ID NO: 11 disclosed in WO 2003/048353, or amino acids -23-353; -23-374; -23-397; 1-353; 1-374; 1-397; 177-353; 177-374; or 177-397 thereof, and SEQ ID NO: 10 disclosed in WO 2003/048353.

Another metalloprotease suitable for use in a process of the invention is the *Aspergillus oryzae* metalloprotease comprising SEQ ID NO: 5 of WO 2010/008841, or a metalloprotease is an isolated polypeptide which has a degree of identity to SEQ ID NO: 5 of at least about 80%, at least 82%, at least 85%, at least 90%, at least 95%, or at least 97%; and which have metalloprotease activity. In particular embodiments, the metalloprotease consists of the amino acid sequence of SEQ ID NO: 5 of WO 2010/008841.

In a particular embodiment, a metalloprotease has an amino acid sequence that differs by forty, thirty-five, thirty, twenty-five, twenty, or by fifteen amino acids from amino acids -178 to 177, -159 to 177, or +1 to 177 of the amino acid sequences of the *Thermoascus aurantiacus* or *Aspergillus oryzae* metalloprotease.

In another embodiment, a metalloprotease has an amino acid sequence that differs by ten, or by nine, or by eight, or by seven, or by six, or by five amino acids from amino acids -178 to 177, -159 to 177, or +1 to 177 of the amino acid sequences of these metalloproteases, e.g., by four, by three, by two, or by one amino acid.

In particular embodiments, the metalloprotease a) comprises or b) consists of

i) the amino acid sequence of amino acids -178 to 177, -159 to 177, or +1 to 177 of SEQ ID NO:1 of WO 2010/008841;

ii) the amino acid sequence of amino acids -23-353, -23-374, -23-397, 1-353, 1-374, 1-397, 177-353, 177-374, or 177-397 of SEQ ID NO: 3 of WO 2010/008841;

iii) the amino acid sequence of SEQ ID NO: 5 of WO 2010/008841; or

allelic variants, or fragments, of the sequences of i), ii), and iii) that have protease activity.

A fragment of amino acids -178 to 177, -159 to 177, or +1 to 177 of SEQ ID NO: 1 of WO 2010/008841 or of amino acids -23-353, -23-374, -23-397, 1-353, 1-374, 1-397, 177-353, 177-374, or 177-397 of SEQ ID NO: 3 of WO 2010/008841; is a polypeptide having one or more amino acids deleted from the amino and/or carboxyl terminus of these amino acid sequences. In one embodiment a fragment contains at least 75 amino acid residues, or at least 100 amino acid residues, or at least 125 amino acid residues, or at least 150 amino acid residues, or at least 160 amino acid residues, or at least 165 amino acid residues, or at least 170 amino acid residues, or at least 175 amino acid residues.

CA 03158982 2022-5-19

To determine whether a given protease is a metallo protease or not, reference is made to the above "Handbook of Proteolytic Enzymes" and the principles indicated therein. Such determination can be carried out for all types of proteases, be it naturally occurring or wild-type proteases; or genetically engineered or synthetic proteases.

The protease may be a variant of, e.g., a wild-type protease, having thermostability properties defined herein. In one embodiment, the thermostable protease is a variant of a metallo protease. In one embodiment, the thermostable protease used in a process described herein is of fungal origin, such as a fungal metallo protease, such as a fungal metallo protease derived from a strain of the genus *Thermoascus*, preferably a strain of *Thermoascus aurantiacus*, especially *Thermoascus aurantiacus* CGMCC No. 0670 (classified as EC 3.4.24.39).

In one embodiment, the thermostable protease is a variant of the mature part of the metallo protease shown in SEQ ID NO: 2 disclosed in WO 2003/048353 or the mature part of SEQ ID NO: 1 in WO 2010/008841 further with one of the following substitutions or combinations of substitutions:

S5*+D79L+S87P+A112P+D142L;
D79L+S87P+A112P+T124V+D142L;
S5*+N26R+D79L+S87P+A112P+D142L;
N26R+T46R+D79L+S87P+A112P+D142L;
T46R+D79L+S87P+T116V+D142L;
D79L+P81R+S87P+A112P+D142L;
A27K+D79L+S87P+A112P+T124V+D142L;
D79L+Y82F+S87P+A112P+T124V+D142L;
D79L+Y82F+S87P+A112P+T124V+D142L;
D79L+S87P+A112P+T124V+A126V+D142L;
D79L+S87P+A112P+D142L;
D79L+Y82F+S87P+A112P+D142L;
S38T+D79L+S87P+A112P+A126V+D142L;
D79L+Y82F+S87P+A112P+A126V+D142L;
A27K+D79L+S87P+A112P+A126V+D142L;
D79L+S87P+N98C+A112P+G135C+D142L;
D79L+S87P+A112P+D142L+T141C+M161C;
S36P+D79L+S87P+A112P+D142L;
A37P+D79L+S87P+A112P+D142L;
S49P+D79L+S87P+A112P+D142L;

CA 03158982 2022-5-19

S50P+D79L+S87P+A112P+D142L;
D79L+S87P+D104P+A112P+D142L;
D79L+Y82F+S87G+A112P+D142L;
S70V+D79L+Y82F+S87G+Y97W+A112P+D142L;
D79L+Y82F+S87G+Y97W+D104P+A112P+D142L;
S70V+D79L+Y82F+S87G+A112P+D142L;
D79L+Y82F+S87G+D104P+A112P+D142L;
D79L+Y82F+S87G+A112P+A126V+D142L;
Y82F+S87G+S70V+D79L+D104P+A112P+D142L;
Y82F+S87G+D79L+D104P+A112P+A126V+D142L;
A27K+D79L+Y82F+S87G+D104P+A112P+A126V+D142L;
A27K+Y82F+S87G+D104P+A112P+A126V+D142L;
A27K+D79L+Y82F+ D104P+A112P+A126V+D142L;
A27K+Y82F+D104P+A112P+A126V+D142L;
A27K+D79L+S87P+A112P+D142L; and
D79L+S87P+D142L.

In one embodiment, the thermostable protease is a variant of the metallo protease disclosed as the mature part of SEQ ID NO: 2 disclosed in WO 2003/048353 or the mature part of SEQ ID NO: 1 in WO 2010/008841 with one of the following substitutions or combinations of substitutions:

D79L+S87P+A112P+D142L;
D79L+S87P+D142L; and
A27K+ D79L+Y82F+S87G+D104P+A112P+A126V+D142L.

In one embodiment, the protease variant has at least 75% identity preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, even more preferably at least 93%, most preferably at least 94%, and even most preferably at least 95%, such as even at least 96%, at least 97%, at least 98%, at least 99%, but less than 100% identity to the mature part of the polypeptide of SEQ ID NO: 2 disclosed in WO 2003/048353 or the mature part of SEQ ID NO: 1 in WO 2010/008841.

The thermostable protease may also be derived from any bacterium as long as the protease has the thermostability properties.

In one embodiment, the thermostable protease is derived from a strain of the bacterium *Pyrococcus*, such as a strain of *Pyrococcus furiosus* (pfu protease).

In one embodiment, the protease is one shown as SEQ ID NO: 1 in US patent No. 6,358,726-B1 (Takara Shuzo Company).

In one embodiment, the thermostable protease is a protease having a mature polypeptide sequence of at least 80% identity, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identity to SEQ ID NO: 1 in US patent no. 6,358,726-B1. The *Pyroccus furiosus* protease can be purchased from Takara Bio, Japan.

The *Pyrococcus furiosus* protease may be a thermostable protease as described in SEQ ID NO: 13 of WO2018/098381. This protease (PfuS) was found to have a thermostability of 110% (80°C/70°C) and 103% (90°C/70°C) at pH 4.5 determined.

In one embodiment a thermostable protease used in a process described herein has a thermostability value of more than 20% determined as Relative Activity at 80ºC/70ºC determined as described in Example 2 of WO2018/098381.

In one embodiment, the protease has a thermostability of more than 30%, more than 40%, more than 50%, more than 60%, more than 70%, more than 80%, more than 90%, more than 100%, such as more than 105%, such as more than 110%, such as more than 115%, such as more than 120% determined as Relative Activity at 80ºC/70ºC.

In one embodiment, protease has a thermostability of between 20 and 50%, such as between 20 and 40%, such as 20 and 30% determined as Relative Activity at 80ºC/70ºC. In one embodiment, the protease has a thermostability between 50 and 115%, such as between 50 and 70%, such as between 50 and 60%, such as between 100 and 120%, such as between 105 and 115% determined as Relative Activity at 80ºC/70ºC.

In one embodiment, the protease has a thermostability value of more than 10% determined as Relative Activity at 85ºC/70ºC determined as described in Example 2 of WO2018/098381.

In one embodiment, the protease has a thermostability of more than 10%, such as more than 12%, more than 14%, more than 16%, more than 18%, more than 20%, more than 30%, more than 40%, more that 50%, more than 60%, more than 70%, more than 80%, more than 90%, more than 100%, more than 110% determined as Relative Activity at 85ºC/70ºC.

In one embodiment, the protease has a thermostability of between 10% and 50%, such as between 10% and 30%, such as between 10% and 25% determined as Relative Activity at 85ºC/70ºC.

In one embodiment, the protease has more than 20%, more than 30%, more than 40%, more than 50%, more than 60%, more than 70%, more than 80%, more than 90% determined as

Remaining Activity at 80ºC; and/or the protease has more than 20%, more than 30%, more than 40%, more than 50%, more than 60%, more than 70%, more than 80%, more than 90% determined as Remaining Activity at 84ºC.

Determination of "Relative Activity" and "Remaining Activity" is done as described in Example 2 of WO2018/098381.

In one embodiment, the protease may have a thermostability for above 90, such as above 100 at 85ºC as determined using the Zein-BCA assay as disclosed in Example 3 of WO2018/098381.

In one embodiment, the protease has a thermostability above 60%, such as above 90%, such as above 100%, such as above 110% at 85ºC as determined using the Zein-BCA assay of WO2018/098381.

In one embodiment, protease has a thermostability between 60-120, such as between 70-120%, such as between 80-120%, such as between 90-120%, such as between 100-120%, such as 110-120% at 85ºC as determined using the Zein-BCA assay of WO2018/098381.

In one embodiment, the thermostable protease has at least 20%, such as at least 30%, such as at least 40%, such as at least 50%, such as at least 60%, such as at least 70%, such as at least 80%, such as at least 90%, such as at least 95%, such as at least 100% of the activity of the JTP196 protease variant or Protease Pfu determined by the AZCL-casein assay of WO2018/098381, and described herein.

In one embodiment, the thermostable protease has at least 20%, such as at least 30%, such as at least 40%, such as at least 50%, such as at least 60%, such as at least 70%, such as at least 80%, such as at least 90%, such as at least 95%, such as at least 100% of the protease activity of the Protease 196 variant or Protease Pfu determined by the AZCL-casein assay of WO2018/098381.

Pullulanases

The host cells and fermenting organisms may express a heterologous pullulanase. The pullulanase can be any protease that is suitable for the host cells and fermenting organisms and/or their methods of use described herein, such as a naturally occurring pullulanase or a variant thereof that retains pullulanase activity. Any pullulanase contemplated for expression by a host cell or fermenting organism described below is also contemplated for embodiments of the invention involving exogenous addition of a pullulanase (e.g., added before, during or after liquefaction and/or saccharification).

In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding a pullulanase has an increased level of pullulanase activity compared to the host cells without the heterologous polynucleotide encoding the pullulanase, when cultivated under the same conditions. In some embodiments, the host cell or fermenting organism has an increased level of pullulanase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the pullulanase, when cultivated under the same conditions.

Exemplary pullulanases that can be used with the host cells and/or the methods described herein include bacterial, yeast, or filamentous fungal pullulanases, e.g., obtained from any of the microorganisms described or referenced herein.

Contemplated pullulanases include the pullulanases from *Bacillus amyloderamificans* disclosed in U.S. Patent No. 4,560,651 (hereby incorporated by reference), the pullulanase disclosed as SEQ ID NO: 2 in WO 01/151620 (hereby incorporated by reference), the *Bacillus deramificans* disclosed as SEQ ID NO: 4 in WO 01/151620 (hereby incorporated by reference), and the pullulanase from *Bacillus acidopullulyticus* disclosed as SEQ ID NO: 6 in WO 01/151620 (hereby incorporated by reference) and also described in FEMS Mic. Let. (1994) 115, 97-106.

Additional pullulanases contemplated include the pullulanases from *Pyrococcus woesei*, specifically from *Pyrococcus woesei* DSM No. 3773 disclosed in WO92/02614.

In one embodiment, the pullulanase is a family GH57 pullulanase. In one embodiment, the pullulanase includes an X47 domain as disclosed in US 61/289,040 published as WO 2011/087836 (which are hereby incorporated by reference). More specifically the pullulanase may be derived from a strain of the genus *Thermococcus*, including *Thermococcus litoralis* and *Thermococcus hydrothermalis*, such as the *Thermococcus hydrothermalis* pullulanase truncated at site X4 right after the X47 domain (i.e., amino acids 1-782). The pullulanase may also be a hybrid of the *Thermococcus litoralis* and *Thermococcus hydrothermalis* pullulanases or a *T. hydrothermalis/T. litoralis* hybrid enzyme with truncation site X4 disclosed in US 61/289,040 published as WO 2011/087836 (which is hereby incorporated by reference).

In another embodiment, the pullulanase is one comprising an X46 domain disclosed in WO 2011/076123 (Novozymes).

The pullulanase may be added in an effective amount which include the preferred amount of about 0.0001-10 mg enzyme protein per gram DS, preferably 0.0001-0.10 mg enzyme protein per gram DS, more preferably 0.0001-0.010 mg enzyme protein per gram DS. Pullulanase activity

CA 03158982 2022-5-19

may be determined as NPUN. An Assay for determination of NPUN is described in WO2018/098381.

Suitable commercially available pullulanase products include PROMOZYME D, PROMOZYME™ D2 (Novozymes A/S, Denmark), OPTIMAX L-300 (DuPont-Danisco, USA), and AMANO 8 (Amano, Japan).

In one embodiment, the pullulanase is derived from the *Bacillus subtilis* pullulanase of SEQ ID NO: 114. In one embodiment, the pullulanase is derived from the *Bacillus licheniformis* pullulanase of SEQ ID NO: 115. In one embodiment, the pullulanase is derived from the *Oryza sativa* pullulanase of SEQ ID NO: 116. In one embodiment, the pullulanase is derived from the *Triticum aestivum* pullulanase of SEQ ID NO: 117. In one embodiment, the pullulanase is derived from the *Clostridium phytofermentans* pullulanase of SEQ ID NO: 118. In one embodiment, the pullulanase is derived from the *Streptomyces avermitilis* pullulanase of SEQ ID NO: 119. In one embodiment, the pullulanase is derived from the *Klebsiella pneumoniae* pullulanase of SEQ ID NO: 120.

Additional pullulanases contemplated for use with the present invention can be found in WO2011/153516 (the content of which is incorporated herein).

Additional polynucleotides encoding suitable pullulanases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org).

The pullulanase coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding pullulanases from strains of different genera or species, as described *supra*.

The polynucleotides encoding pullulanases may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding pullulanases are described *supra*.

In one embodiment, the pullulanase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any one of the pullulanases described or referenced herein (e.g., any one of SEQ ID NOs: 114-120). In another embodiment, the pullulanase has a mature polypeptide sequence that is a fragment of the any one of the pullulanases described or referenced herein (e.g., any one of SEQ ID NOs: 114-120). In one embodiment, the number of amino acid residues in the fragment is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the

CA 03158982 2022-5-19

number of amino acid residues in referenced full length pullulanase. In other embodiments, the pullulanase may comprise the catalytic domain of any pullulanase described or referenced herein (e.g., any one of SEQ ID NOs: 114-120).

The pullulanase may be a variant of any one of the pullulanases described *supra* (e.g., any one of SEQ ID NOs: 114-120). In one embodiment, the pullulanase has a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of the pullulanases described *supra* (e.g., any one of SEQ ID NOs: 114-120).

Examples of suitable amino acid changes, such as conservative substitutions that do not significantly affect the folding and/or activity of the pullulanase, are described herein.

In one embodiment, the pullulanase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from the amino acid sequence of any one of the pullulanases described *supra* (e.g., any one of SEQ ID NOs: 114-120). In one embodiment, the pullulanase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) of amino acid sequence of any one of the pullulanases described *supra* (e.g., any one of SEQ ID NOs: 114-120). In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the pullulanase has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the pullulanase activity of any pullulanase described or referenced herein under the same conditions (e.g., any one of SEQ ID NOs: 114-120).

In one embodiment, the pullulanase coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any pullulanase described or referenced herein (e.g., any one of SEQ ID NOs: 114-120). In one embodiment, the pullulanase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any pullulanase described or referenced herein (e.g., any one of SEQ ID NOs: 114-120).

In one embodiment, the pullulanase comprises the coding sequence of any pullulanase described or referenced herein (e.g., any one of SEQ ID NOs: 114-120). In one embodiment, the

CA 03158982 2022-5-19

pullulanase comprises a coding sequence that is a subsequence of the coding sequence from any pullulanase described or referenced herein, wherein the subsequence encodes a polypeptide having pullulanase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The referenced pullulanase coding sequence of any related aspect or embodiment described herein can be the native coding sequence or a degenerate sequence, such as a codon-optimized coding sequence designed for use in a particular host cell (e.g., optimized for expression in *Saccharomyces cerevisiae*).

The pullulanase can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

Gene Disruptions

The host cells and fermenting organisms described herein may also comprise one or more (e.g., two, several) gene disruptions, e.g., to divert sugar metabolism from undesired products to ethanol. In some embodiments, the recombinant host cells produce a greater amount of ethanol compared to the cell without the one or more disruptions when cultivated under identical conditions. In some embodiments, one or more of the disrupted endogenous genes is inactivated. In some embodiments, the host cell or fermenting organism is a diploid and has a disruption (e.g., inactivation) of both copies of the referenced gene.

In certain embodiments, the host cell or fermenting organism provided herein comprises a disruption of one or more endogenous genes encoding enzymes involved in producing alternate fermentative products such as glycerol or other byproducts such as acetate or diols. For example, the cells provided herein may comprise a disruption of one or more endogenous genes encoding a glycerol 3-phosphatase (GPP, E.C. 3.1.3.21, catalyzes conversion of glycerol-3 phosphate to glycerol), a glycerol 3-phosphate dehydrogenase (GPD, catalyzes reaction of dihydroxyacetone phosphate to glycerol 3-phosphate), glycerol kinase (catalyzes conversion of glycerol 3-phosphate to glycerol), dihydroxyacetone kinase (catalyzes conversion of dihydroxyacetone phosphate to dihydroxyacetone), glycerol dehydrogenase (catalyzes conversion of dihydroxyacetone to glycerol), and aldehyde dehydrogenase (ALD, e.g., converts acetaldehyde to acetate).

In some embodiments, the host cell or fermenting organism comprises a disruption to one or more endogenous genes encoding a glycerol 3-phosphatase (GPP). *Saccharomyces cerevisiae* has two glycerol-3-phosphate phosphatase paralogs encoding GPP1 (UniProt No.

CA 03158982 2022-5-19

P41277; SEQ ID NO: 402) and GPP2 (UniProt No. P40106; SEQ ID NO: 403) (Pahlman et al. (2001) *J. Biol. Chem.* 276(5):3555-63; Norbeck et al. (1996) *J. Biol. Chem.* 271(23):13875-81). In some embodiments, the host cell or fermenting organism comprises a disruption to GPP1. In some embodiments, the host cell or fermenting organism comprises a disruption to GPP2. In some embodiments, the host cell or fermenting organism comprises a disruption to GPP1 and GPP2.

In some embodiments, the host cell or fermenting organism comprises a disruption to one or more endogenous genes encoding a glycerol 3-phosphate dehydrogenase (GPD). *Saccharomyces cerevisiae* has two glycerol 3-phosphate dehydrogenases which encode GPD1 (UniProt No. Q00055; SEQ ID NO: 404) and GPD2 (UniProt No. P41911; SEQ ID NO: 405). In some embodiments, the host cell or fermenting organism comprises a disruption to GPD1. In some embodiments, the host cell or fermenting organism comprises a disruption to GPD2. In some embodiments, the host cell or fermenting organism comprises a disruption to GPD1 and GPD2.

In some embodiments, the host cell or fermenting organism comprises a disruption to an endogenous gene encoding GPP (e.g., GPP1 and/or GPP2) and/or a GPD (GPD1 and/or GPD2), wherein the host cell or fermenting organism produces a decreased amount of glycerol (e.g., at least 25% less, at least 50% less, at least 60% less, at least 70% less, at least 80% less, or at least 90% less) compared to the cell without the disruption to the endogenous gene encoding the GPP and/or GPD when cultivated under identical conditions.

Modeling analysis can be used to design gene disruptions that additionally optimize utilization of the pathway. One exemplary computational method for identifying and designing metabolic alterations favoring biosynthesis of a desired product is the OptKnock computational framework, Burgard *et al.*, 2003, *Biotechnol. Bioeng.* 84: 647-657.

The host cells and fermenting organisms comprising a gene disruption may be constructed using methods well known in the art, including those methods described herein. A portion of the gene can be disrupted such as the coding region or a control sequence required for expression of the coding region. Such a control sequence of the gene may be a promoter sequence or a functional part thereof, *i.e.*, a part that is sufficient for affecting expression of the gene. For example, a promoter sequence may be inactivated resulting in no expression or a weaker promoter may be substituted for the native promoter sequence to reduce expression of the coding sequence. Other control sequences for possible modification include, but are not limited to, a leader, propeptide sequence, signal sequence, transcription terminator, and transcriptional activator.

CA 03158982 2022-5-19

The host cells and fermenting organisms comprising a gene disruption may be constructed by gene deletion techniques to eliminate or reduce expression of the gene. Gene deletion techniques enable the partial or complete removal of the gene thereby eliminating their expression. In such methods, deletion of the gene is accomplished by homologous recombination using a plasmid that has been constructed to contiguously contain the 5' and 3' regions flanking the gene.

The host cells and fermenting organisms comprising a gene disruption may also be constructed by introducing, substituting, and/or removing one or more (e.g., two, several) nucleotides in the gene or a control sequence thereof required for the transcription or translation thereof. For example, nucleotides may be inserted or removed for the introduction of a stop codon, the removal of the start codon, or a frame-shift of the open reading frame. Such a modification may be accomplished by site-directed mutagenesis or PCR generated mutagenesis in accordance with methods known in the art. See, for example, Botstein and Shortle, 1985, *Science* 229: 4719; Lo *et al.*, 1985, *Proc. Natl. Acad. Sci. U.S.A.* 81: 2285; Higuchi *et al.*, 1988, *Nucleic Acids Res* 16: 7351; Shimada, 1996, *Meth. Mol. Biol.* 57: 157; Ho *et al.*, 1989, *Gene* 77: 61; Horton *et al.*, 1989, *Gene* 77: 61; and Sarkar and Sommer, 1990, *BioTechniques* 8: 404.

The host cells and fermenting organisms comprising a gene disruption may also be constructed by inserting into the gene a disruptive nucleic acid construct comprising a nucleic acid fragment homologous to the gene that will create a duplication of the region of homology and incorporate construct DNA between the duplicated regions. Such a gene disruption can eliminate gene expression if the inserted construct separates the promoter of the gene from the coding region or interrupts the coding sequence such that a non-functional gene product results. A disrupting construct may be simply a selectable marker gene accompanied by 5' and 3' regions homologous to the gene. The selectable marker enables identification of transformants containing the disrupted gene.

The host cells and fermenting organisms comprising a gene disruption may also be constructed by the process of gene conversion (see, for example, Iglesias and Trautner, 1983, *Molecular General Genetics* 189: 73-76). For example, in the gene conversion method, a nucleotide sequence corresponding to the gene is mutagenized *in vitro* to produce a defective nucleotide sequence, which is then transformed into the recombinant strain to produce a defective gene. By homologous recombination, the defective nucleotide sequence replaces the endogenous gene. It may be desirable that the defective nucleotide sequence also comprises a marker for selection of transformants containing the defective gene.

CA 03158982 2022-5-19

The host cells and fermenting organisms comprising a gene disruption may be further constructed by random or specific mutagenesis using methods well known in the art, including, but not limited to, chemical mutagenesis (see, for example, Hopwood, *The Isolation of Mutants* in *Methods in Microbiology* (J.R. Norris and D.W. Ribbons, eds.) pp. 363-433, Academic Press, New York, 1970). Modification of the gene may be performed by subjecting the parent strain to mutagenesis and screening for mutant strains in which expression of the gene has been reduced or inactivated. The mutagenesis, which may be specific or random, may be performed, for example, by use of a suitable physical or chemical mutagenizing agent, use of a suitable oligonucleotide, or subjecting the DNA sequence to PCR generated mutagenesis. Furthermore, the mutagenesis may be performed by use of any combination of these mutagenizing methods.

Examples of a physical or chemical mutagenizing agent suitable for the present purpose include ultraviolet (UV) irradiation, hydroxylamine, N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), N-methyl-N'-nitrosogaunidine (NTG) O-methyl hydroxylamine, nitrous acid, ethyl methane sulphonate (EMS), sodium bisulphite, formic acid, and nucleotide analogues. When such agents are used, the mutagenesis is typically performed by incubating the parent strain to be mutagenized in the presence of the mutagenizing agent of choice under suitable conditions, and selecting for mutants exhibiting reduced or no expression of the gene.

A nucleotide sequence homologous or complementary to a gene described herein may be used from other microbial sources to disrupt the corresponding gene in a recombinant strain of choice.

In one embodiment, the modification of a gene in the recombinant cell is unmarked with a selectable marker. Removal of the selectable marker gene may be accomplished by culturing the mutants on a counter-selection medium. Where the selectable marker gene contains repeats flanking its 5' and 3' ends, the repeats will facilitate the looping out of the selectable marker gene by homologous recombination when the mutant strain is submitted to counter-selection. The selectable marker gene may also be removed by homologous recombination by introducing into the mutant strain a nucleic acid fragment comprising 5' and 3' regions of the defective gene, but lacking the selectable marker gene, followed by selecting on the counter-selection medium. By homologous recombination, the defective gene containing the selectable marker gene is replaced with the nucleic acid fragment lacking the selectable marker gene. Other methods known in the art may also be used.

CA 03158982 2022-5-19

Active pentose fermenation pathway

The host cells or fermenting organisms described herein (e.g., yeast cells) may comprise an active pentose fermentation pathway, such as an active xylose fermentation pathway and/or and active arabinose fermentation pathway as described in more detail below. Pentose fermentation pathways and pathway genes and corresponding engineered transformants for fermentation of pentose (e.g., xylose, arabinose) are known in the art.

Any suitable pentose fermentation pathway gene, endogenous or heterologous, may be used and expressed in sufficient amount to produce an enzyme involved in a selected pentose fermentation pathway. With the complete genome sequence available for now numerous microorganism genomes and a variety of yeast, fungi, plant, and mammalian genomes, the identification of genes encoding the selected pentose fermentation pathway enzymatic activities taught herein is routine and well known in the art for a selected host. For example, suitable homologues, orthologs, paralogs and nonorthologous gene displacements of known genes, and the interchange of genetic alterations between organisms can be identified in related or distant host to a selected host.

For host cells without a known genome sequence, sequences for genes of interest (either as overexpression candidates or as insertion sites) can typically be obtained using techniques known in the art. Routine experimental design can be employed to test expression of various genes and activity of various enzymes, including genes and enzymes that function in a pentose fermentation pathway. Experiments may be conducted wherein each enzyme is expressed in the cell individually and in blocks of enzymes up to and including preferably all pathway enzymes, to establish which are needed (or desired) for improved pentose fermentation. One illustrative experimental design tests expression of each individual enzyme as well as of each unique pair of enzymes, and further can test expression of all required enzymes, or each unique combination of enzymes. A number of approaches can be taken, as will be appreciated.

The host cells of the invention can be produced by introducing heterologous polynucleotides encoding one or more of the enzymes participating in an active pentose fermentation pathway, as described below. As one in the art will appreciate, in some instances (e.g., depending on the selection of host) the heterologous expression of every gene shown in the active pentose fermentation may not be required since a host cell may have endogenous enzymatic activity from one or more pathway genes. For example, if a chosen host is deficient in one or more enzymes of an active pentose fermentation pathway, then heterologous polynucleotides for the deficient enzyme(s) are introduced into the host for subsequent expression. Alternatively, if the chosen host exhibits endogenous expression of some pathway

CA 03158982 2022-5-19

genes, but is deficient in others, then an encoding polynucleotide is needed for the deficient enzyme(s) to achieve pentose fermentation. Thus, a recombinant host cell of the invention can be produced by introducing heterologous polynucleotides to obtain the enzyme activities of a desired biosynthetic pathway or a desired biosynthetic pathway can be obtained by introducing one or more heterologous polynucleotides that, together with one or more endogenous enzymes, produces a desired product such as ethanol.

Depending on the pentose fermentation pathway constituents of a selected recombinant host organism, the host cells of the invention will include at least one heterologous polynucleotide and optionally up to all encoding heterologous polynucleotides for the pentose fermentation pathway. For example, pentose fermentation can be established in a host deficient in a pentose fermentation pathway enzyme through heterologous expression of the corresponding polynucleotide. In a host deficient in all enzymes of a pentose fermentation pathway, heterologous expression of all enzymes in the pathway can be included, although it is understood that all enzymes of a pathway can be expressed even if the host contains at least one of the pathway enzymes.

The enzymes of the selected active pentose fermentation pathway, and activities thereof, can be detected using methods known in the art or as described herein. These detection methods may include use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme substrate. *See*, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual,* Third Ed., Cold Spring Harbor Laboratory, New York (2001); Ausubel et al., *Current Protocols in Molecular Biology,* John Wiley and Sons, Baltimore, MD (1999); and Hanai et al., *Appl. Environ. Microbiol.* 73:7814-7818 (2007)).

The active pentose fermentation pathway may be an active xylose fermentation pathway. Exemplary xylose fermentation pathways are known in the art (e.g., WO2003/062430, WO2003/078643, WO2004/067760, WO2006/096130, WO2009/017441, WO2010/059095, WO2011/059329, WO2011/123715, WO2012/113120, WO2012/135110, WO2013/081700, WO2018/112638 and US2017/088866). Any xylose fermentation pathway or gene thereof described in the foregoing references is incorporated herein by reference for use in Applicant's active xylose fermentation pathway. Figure 2 shows conversion of D-xylose to D-xylulose 5-phosphate, which is then fermented to ethanol via the pentose phosphate pathway. The oxido-reductase pathway uses an aldolase reductase (AR, such as xylose reductase (XR)) to reduce D-xylose to xylitol followed by oxidation of xylitol to D-xylulose with xylitol dehydrogenase (XDH; also known as D-xylulose reductase). The isomerase pathway uses xylose isomerase (XI) to

CA 03158982 2022-5-19

convert D-xylose into D-xylulose. D-xylulose is then converted to D-xylulose-5-phosphate with xylulokinase (XK)

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding a xylose isomerase (XI). The xylose isomerase may be any xylose isomerase that is suitable for the host cells and the methods described herein, such as a naturally occurring xylose isomerase or a variant thereof that retains xylose isomerase activity. In one embodiment, the xylose isomerase is present in the cytosol of the host cells.

In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding a xylose isomerase has an increased level of xylose isomerase activity compared to the host cells without the heterologous polynucleotide encoding the xylose isomerase, when cultivated under the same conditions. In some embodiments, the host cells or fermenting organisms have an increased level of xylose isomerase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cells without the heterologous polynucleotide encoding the xylose isomerase, when cultivated under the same conditions.

Exemplary xylose isomerases that can be used with the recombinant host cells and methods of use described herein include, but are not limited to, XIs from the fungus *Piromyces sp.* (WO2003/062430) or other sources (Madhavan et al., 2009, *Appl Microbiol Biotechnol.* 82(6), 1067-1078) have been expressed in *S. cerevisiae* host cells. Still other XIs suitable for expression in yeast have been described in US 2012/0184020 (an XI from *Ruminococcus flavefaciens*), WO2011/078262 (several XIs from *Reticulitermes speratus* and *Mastotermes darwiniensis*) and WO2012/009272 (constructs and fungal cells containing an XI from *Abiotrophia defectiva*). US 8,586,336 describes a *S. cerevisiae* host cell expressing an XI obtained by bovine rumen fluid (shown herein as SEQ ID NO: 74).

Additional polynucleotides encoding suitable xylose isomerases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org). In one embodiment, the xylose isomerases is a bacterial, a yeast, or a filamentous fungal xylose isomerase, e.g., obtained from any of the microorganisms described or referenced herein, as described *supra*.

The xylose isomerase coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding xylose isomerases from strains of different genera or species, as described *supra*.

CA 03158982 2022-5-19

The polynucleotides encoding xylose isomerases may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding xylose isomerases are described *supra*.

In one embodiment, the xylose isomerase has a mature polypeptide sequence of having at least 60%, e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any xylose isomerase described or referenced herein (e.g., the xylose isomerase of SEQ ID NO: 74). In one embodiment, the xylose isomerase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any xylose isomerase described or referenced herein (e.g., the xylose isomerase of SEQ ID NO: 74). In one embodiment, the xylose isomerase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any xylose isomerase described or referenced herein (e.g., the xylose isomerase of SEQ ID NO: 74), allelic variant, or a fragment thereof having xylose isomerase activity. In one embodiment, the xylose isomerase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) amino acids. In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the xylose isomerase has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the xylose isomerase activity of any xylose isomerase described or referenced herein (e.g., the xylose isomerase of SEQ ID NO: 74) under the same conditions.

In one embodiment, the xylose isomerase coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any xylose isomerase described or referenced herein (e.g., the xylose isomerase of SEQ ID NO: 74). In one embodiment, the xylose isomerase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least

CA 03158982 2022-5-19

96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any xylose isomerase described or referenced herein (e.g., the xylose isomerase of SEQ ID NO: 74).

In one embodiment, the heterologous polynucleotide encoding the xylose isomerase comprises the coding sequence of any xylose isomerase described or referenced herein (e.g., the xylose isomerase of SEQ ID NO: 74). In one embodiment, the heterologous polynucleotide encoding the xylose isomerase comprises a subsequence of the coding sequence from any xylose isomerase described or referenced herein, wherein the subsequence encodes a polypeptide having xylose isomerase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The xylose isomerases can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

The host cell or fermenting organism may also comprise an aldose reductase (AR), xylitol dehydrogenase (XDH) and/or xylulokinase (XK) as described below.

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding a ribulose 5 phosphate 3-epimerase (RPE1). A ribulose 5 phosphate 3-epimerase, as used herein, provides enzymatic activity for converting L-ribulose 5-phosphate to L-xylulose 5-phosphate (EC 5.1.3.22). The RPE1 may be any RPE1 that is suitable for the host cells and the methods described herein, such as a naturally occurring RPE1 or a variant thereof that retains RPE1 activity. In one embodiment, the RPE1 is present in the cytosol of the host cells.

In one embodiment, the recombinant cell comprises a heterologous polynucleotide encoding a ribulose 5 phosphate 3-epimerase (RPE1), wherein the RPE1 is *Saccharomyces cerevisiae* RPE1, or an RPE1 having at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to a *Saccharomyces cerevisiae* RPE1.

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding a ribulose 5 phosphate isomerase (RKI1). A ribulose 5 phosphate isomerase, as used herein, provides enzymatic activity for converting ribose-5-phophate to ribulose 5-phosphate. The RKI1 may be any RKI1 that is suitable for the host cells and the methods described herein, such as a naturally occurring RKI1 or a variant thereof that retains RKI1 activity. In one embodiment, the RKI1 is present in the cytosol of the host cells.

In one embodiment, the host cell or fermenting organism comprises a heterologous polynucleotide encoding a ribulose 5 phosphate isomerase (RKI1), wherein the RKI1 is a *Saccharomyces cerevisiae* RKI1, or an RKI1 having a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to a *Saccharomyces cerevisiae* RKI1.

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding a transketolase (TKL1). The TKL1 may be any TKL1 that is suitable for the host cells and the methods described herein, such as a naturally occurring TKL1 or a variant thereof that retains TKL1 activity. In one embodiment, the TKL1 is present in the cytosol of the host cells.

In one embodiment, the host cell or fermenting organism comprises a heterologous polynucleotide encoding a transketolase (TKL1), wherein the TKL1 is a *Saccharomyces cerevisiae* TKL1, or a TKL1 having a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to a *Saccharomyces cerevisiae* TKL1.

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding a transaldolase (TAL1). The TAL1 may be any TAL1 that is suitable for the host cells and the methods described herein, such as a naturally occurring TAL1 or a variant thereof that retains TAL1 activity. In one embodiment, the TAL1 is present in the cytosol of the host cells.

In one embodiment, the host cell or fermenting organism comprises a heterologous polynucleotide encoding a transketolase (TAL1), wherein the TAL1 is a *Saccharomyces cerevisiae* TAL1, or a TAL1 having a mature polypeptide sequence of at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to a *Saccharomyces cerevisiae* TAL1.

The active pentose fermentation pathway may be an active arabinose fermentation pathway. Exemplary arabinose fermentation pathways are known in the art (e.g., WO2002/066616; WO2003/095627; WO2007/143245; WO2008/041840; WO2009/011591; WO2010/151548; WO2011/003893; WO2011/131674; WO2012/143513; US2012/225464; US 7,977,083). Any arabinose fermentation pathway or gene thereof described in the foregoing references is incorporated herein by reference for use in Applicant's active xylose fermentation pathway. Figure 1 shows arabinose fermentation pathways from L-arabinose to D-xylulose 5-phosphate, which is then fermented to ethanol via the pentose phosphate pathway. The bacterial pathway utilizes genes L-arabinose isomerase (AI, such as *araA*), L-ribulokinase (RK, such as

CA 03158982 2022-5-19

araB), and L-ribulose-5-P4-epimerase (R5PE, such as *araD*) to convert L-arabinose to D-xylulose 5-phosphate. The fungal pathway proceeds using aldose reductase (AR), L-arabinitol 4-dehydrogenase (LAD), L-xylulose reductase (LXR), xylitol dehydrogenase (XDH, also known as D-xylulose reductase) and xylulokinase (XK).

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding a L-xylulose reductase (LXR). As shown in Figure 1, L-xylulose reductase (LXR) is an enzyme used in a "fungal pathway" that proceeds from L-arabinose to D-xyluluose-5-phosphate, where the L-xylulose reductase provides enzymatic activity for converting L-xylulose to xylitol. The L-xylulose reductase may be any L-xylulose reductase that is suitable for the host cells and the methods described herein, such as a naturally occurring L-xylulose reductase or a variant thereof that retains L-xylulose reductase activity. In one embodiment, the L-xylulose reductase is present in the cytosol of the host cells.

In some embodiments, the host cells or fermenting organisms comprising a heterologous polynucleotide encoding an L-xylulose reductase (LXR) have an increased level of L-xylulose reductase activity compared to the host cells without the heterologous polynucleotide encoding the L-xylulose reductase, when cultivated under the same conditions. In some embodiments, the host cells have an increased level of L-xylulose reductase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cells without the heterologous polynucleotide encoding the L-xylulose reductase, when cultivated under the same conditions.

Exemplary L-xylulose reductases (LXRs) that can be used with the host cells and fermenting organisms, and methods of use described herein include, but are not limited to, any one of the L-xylulose reductases of SEQ ID NOs: 454-465, as described in U.S. Provisional Application No. 63/024,010, filed May 13, 2020 (the content of which is hereby incorporated by reference), such as the *A. brasiliensis* L-xylulose reductase of SEQ ID NO: 454, the *T. leycettanus* L-xylulose reductase of SEQ ID NO: 457, the *A. aculeatus* L-xylulose reductase of SEQ ID NO: 459 and the *A. niger* L-xylulose reductase of SEQ ID NO: 461. Additional polynucleotides encoding suitable L-xylulose reductases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org). In one embodiment, the L-xylulose reductase is a bacterial, a yeast, or a filamentous fungal L-xylulose reductase, e.g., obtained from any of the microorganisms described or referenced herein, as described *supra*.

CA 03158982 2022-5-19

The L-xylulose reductase (LXR) coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding L-xylulose reductases from strains of different genera or species, as described *supra*.

The polynucleotides encoding the L-xylulose reductases (LXRs) may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding L-xylulose reductases (LXRs) are described *supra*.

In one embodiment, the L-xylulose reductase (LXR) has a mature polypeptide sequence of at least 60%, e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any L-xylulose reductase described or referenced herein (e.g., any one of L-xylulose reductases of SEQ ID NOs: 454-465, such as the L-xylulose reductase of SEQ ID NO: 454, 457, 459 or 461). In one embodiment, the L-xylulose reductase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any L-xylulose reductase described or referenced herein (e.g., any one of L-xylulose reductases of SEQ ID NOs: 454-465, such as the L-xylulose reductase of SEQ ID NO: 454, 457, 459 or 461). In one embodiment, the L-xylulose reductase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any L-xylulose reductase described or referenced herein (e.g., any one of L-xylulose reductases of SEQ ID NOs: 454-465, such as the L-xylulose reductase of SEQ ID NO: 454, 457, 459 or 461), allelic variant, or a fragment thereof having L-xylulose reductase activity. In one embodiment, the L-xylulose reductase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) amino acids. In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the L-xylulose reductase (LXR) has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the L-xylulose reductase activity of any L-xylulose reductase described or referenced herein (e.g., any one of L-xylulose reductases of SEQ ID NOs: 454-465, such as the L-xylulose reductase of SEQ ID NO: 454, 457, 459 or 461) under the same conditions.

In one embodiment, the L-xylulose reductase (LXR) coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any L-xylulose reductase described or referenced herein (e.g., any one of L-xylulose reductases of SEQ ID NOs: 454-465, such as the L-xylulose reductase of SEQ ID NO: 454, 457, 459 or 461). In one embodiment, the L-xylulose reductase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any L-xylulose reductase described or referenced herein (e.g., any one of L-xylulose reductases of SEQ ID NOs: 454-465, such as the L-xylulose reductase of SEQ ID NO: 454, 457, 459 or 461).

In one embodiment, the heterologous polynucleotide encoding the L-xylulose reductase (LXR) comprises the coding sequence of any L-xylulose reductase described or referenced herein (e.g., any one of L-xylulose reductases of SEQ ID NOs: 454-465, such as the L-xylulose reductase of SEQ ID NO: 454, 457, 459 or 461). In one embodiment, the heterologous polynucleotide encoding the L-xylulose reductase comprises a subsequence of the coding sequence from any L-xylulose reductase described or referenced herein, wherein the subsequence encodes a polypeptide having L-xylulose reductase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The L-xylulose reductases (LXRs) can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding an aldose reductase (AR). An aldose reductase, as used herein, provides enzymatic activity for converting L-arabinose to L-arabitol, and may also have enzymatic activity for converting D-xylose to xylitol (known as a xylose reductase, XR). The aldose reductase may be any aldose reductase that is suitable for the host cells and the methods described herein, such as a naturally occurring aldose reductase or a variant thereof that retains aldose reductase activity. In one embodiment, the aldose reductase is present in the cytosol of the host cells.

In some embodiments, the host cells or fermenting organisms comprising a heterologous polynucleotide encoding an aldose reductase (AR) have an increased level of aldose reductase activity compared to the host cells without the heterologous polynucleotide encoding the aldose

reductase, when cultivated under the same conditions. In some embodiments, the host cells have an increased level of aldose reductase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cells without the heterologous polynucleotide encoding the aldose reductase, when cultivated under the same conditions.

Exemplary aldose reductases (ARs) that can be used with the host cells and fermenting organisms, and methods of use described herein include, but are not limited to, the *Aspergillus niger* aldose reductase of SEQ ID NO: 438, the the *Aspergillus oryzae* aldose reductase of SEQ ID NO: 439, the *Magnaporthe oryzae* aldose reductase of SEQ ID NO: 440, the *Meyerozyma guilliermondii* aldose reductase of SEQ ID NO: 441 and the *Scheffersomyces stipitis* aldose reductase of SEQ ID NO: 442. Additional polynucleotides encoding suitable aldose reductase may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org). In one embodiment, the aldose reductase is a bacterial, a yeast, or a filamentous fungal aldose reductase, e.g., obtained from any of the microorganisms described or referenced herein, as described *supra*.

The aldose reductase (AR) coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding aldose reductases from strains of different genera or species, as described *supra*.

The polynucleotides encoding the aldose reductases (ARs) may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding aldose reductases (ARs) are described *supra*.

In one embodiment, the aldose reductase (AR) has a mature polypeptide sequence of at least 60%, e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any aldose reductase described or referenced herein (e.g., the aldose reductase of SEQ ID NO: 438, 439, 440, 441 or 442). In one embodiment, the aldose reductase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any aldose reductase described or referenced herein (e.g., the aldose reductase of SEQ ID NO: 438, 439, 440, 441 or 442). In one embodiment, the aldose reductase has a mature polypeptide

CA 03158982 2022-5-19

sequence that comprises or consists of the amino acid sequence of any aldose reductase described or referenced herein (e.g., the aldose reductase of SEQ ID NO: 438, 439, 440, 441 or 442), allelic variant, or a fragment thereof having aldose reductase activity. In one embodiment, the aldose reductase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) amino acids. In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the aldose reductase (AR) has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the aldose reductase activity of any aldose reductase described or referenced herein (e.g., the aldose reductase of SEQ ID NO: 438, 439, 440, 441 or 442) under the same conditions.

In one embodiment, the aldose reductase (AR) coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any aldose reductase described or referenced herein (e.g., the aldose reductase of SEQ ID NO: 438, 439, 440, 441 or 442). In one embodiment, the aldose reductase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any aldose reductase described or referenced herein (e.g., the aldose reductase of SEQ ID NO: 438, 439, 440, 441 or 442).

In one embodiment, the heterologous polynucleotide encoding the aldose reductase (AR) comprises the coding sequence of any aldose reductase described or referenced herein (e.g., the aldose reductase of SEQ ID NO: 438, 439, 440, 441 or 442). In one embodiment, the heterologous polynucleotide encoding the aldose reductase comprises a subsequence of the coding sequence from any aldose reductase described or referenced herein, wherein the subsequence encodes a polypeptide having aldose reductase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The aldose reductases (ARs) can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding an L-arabinitol 4-dehydrogenase (LAD). A L-

CA 03158982 2022-5-19

arabinitol 4-dehydrogenase, as used herein, provides enzymatic activity for converting L-arabitol to L-xylulose. The L-arabinitol 4-dehydrogenase may be any L-arabinitol 4-dehydrogenase that is suitable for the host cells and the methods described herein, such as a naturally occurring L-arabinitol 4-dehydrogenase or a variant thereof that retains L-arabinitol 4-dehydrogenase activity. In one embodiment, the L-arabinitol 4-dehydrogenase is present in the cytosol of the host cells.

In some embodiments, the host cells or fermenting organisms comprising a heterologous polynucleotide encoding a L-arabinitol 4-dehydrogenase (LAD) have an increased level of L-arabinitol 4-dehydrogenase activity compared to the host cells without the heterologous polynucleotide encoding the L-arabinitol 4-dehydrogenase, when cultivated under the same conditions. In some embodiments, the host cells have an increased level of L-arabinitol 4-dehydrogenase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cells without the heterologous polynucleotide encoding the L-arabinitol 4-dehydrogenase, when cultivated under the same conditions.

Exemplary L-arabinitol 4-dehydrogenases (LADs) that can be used with the host cells and fermenting organisms, and methods of use described herein include, but are not limited to, the *Meyerozyma caribbica* LAD of SEQ ID NO: 443, the *Trichoderma reesei* LAD of SEQ ID NO: 444, the *Meyerozyma guilliermondii* LAD of SEQ ID NO: 445, the *Candida arabinofermentans* LAD of SEQ ID NO: 446, the *Candida carpophila* LAD of SEQ ID NO: 447, the *Talaromyces emersonii* LAD of SEQ ID NO: 448, the *Aspergillus oryzae* LAD of SEQ ID NO: 449, the *Neurospora crassa* LAD of SEQ ID NO: 450, the *Trichoderma reesei* LAD of SEQ ID NO: 451, the *Aspergillus niger* LAD of SEQ ID NO: 452 and the *Penicillium rubens* LAD of SEQ ID NO: 453. Additional polynucleotides encoding suitable L-arabinitol 4-dehydrogenases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org). In one embodiment, the L-arabinitol 4-dehydrogenase is a bacterial, a yeast, or a filamentous fungal L-arabinitol 4-dehydrogenase, e.g., obtained from any of the microorganisms described or referenced herein, as described *supra*.

The L-arabinitol 4-dehydrogenase (LAD) coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding L-arabinitol 4-dehydrogenases from strains of different genera or species, as described *supra*.

The polynucleotides encoding L-arabinitol 4-dehydrogenases (LADs) may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

CA 03158982 2022-5-19

Techniques used to isolate or clone polynucleotides encoding L-arabinitol 4-dehydrogenases (LADs) are described *supra*.

In one embodiment, the L-arabinitol 4-dehydrogenase (LAD) has a mature polypeptide sequence of at least 60%, e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any L-arabinitol 4-dehydrogenase described or referenced herein (e.g., the L-arabinitol 4-dehydrogenase of SEQ ID NO: 443, 444, 445, 446, 447, 448, 449, 450, 451, 452 or 453). In one embodiment, the L-arabinitol 4-dehydrogenase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any L-arabinitol 4-dehydrogenase described or referenced herein (e.g., the L-arabinitol 4-dehydrogenase of SEQ ID NO: 443, 444, 445, 446, 447, 448, 449, 450, 451, 452 or 453). In one embodiment, the L-arabinitol 4-dehydrogenase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any L-arabinitol 4-dehydrogenase described or referenced herein (e.g., the L-arabinitol 4-dehydrogenase of SEQ ID NO: 443, 444, 445, 446, 447, 448, 449, 450, 451, 452 or 453), allelic variant, or a fragment thereof having L-arabinitol 4-dehydrogenase activity. In one embodiment, the L-arabinitol 4-dehydrogenase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) amino acids. In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the L-arabinitol 4-dehydrogenase (LAD) has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the L-arabinitol 4-dehydrogenase activity of any L-arabinitol 4-dehydrogenase described or referenced herein (e.g., the L-arabinitol 4-dehydrogenase of SEQ ID NO: 443, 444, 445, 446, 447, 448, 449, 450, 451, 452 or 453) under the same conditions.

In one embodiment, the L-arabinitol 4-dehydrogenase (LAD) coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any L-arabinitol 4-dehydrogenase described or referenced herein (e.g., the L-arabinitol 4-dehydrogenase of SEQ ID NO: 443, 444, 445, 446, 447, 448, 449, 450, 451, 452 or 453). In one embodiment, the L-arabinitol 4-dehydrogenase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%,

CA 03158982 2022-5-19

at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any xylulokinase described or referenced herein (e.g., the L-arabinitol 4-dehydrogenase of SEQ ID NO: 443, 444, 445, 446, 447, 448, 449, 450, 451, 452 or 453).

In one embodiment, the heterologous polynucleotide encoding the L-arabinitol 4-dehydrogenase (LAD) comprises the coding sequence of any L-arabinitol 4-dehydrogenase described or referenced herein (e.g., the L-arabinitol 4-dehydrogenase of SEQ ID NO: 443, 444, 445, 446, 447, 448, 449, 450, 451, 452 or 453). In one embodiment, the heterologous polynucleotide encoding the L-arabinitol 4-dehydrogenase comprises a subsequence of the coding sequence from any L-arabinitol 4-dehydrogenase described or referenced herein, wherein the subsequence encodes a polypeptide having L-arabinitol 4-dehydrogenase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The L-arabinitol 4-dehydrogenases (LADs) can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding a xylitol dehydrogenase (XDH). A xylitol dehydrogenase, as used herein, provides enzymatic activity for converting xylitol to D-xylulose. The xylitol dehydrogenase may be any xylitol dehydrogenase that is suitable for the host cells and the methods described herein, such as a naturally occurring xylitol dehydrogenase or a variant thereof that retains xylitol dehydrogenase activity. In one embodiment, the xylitol dehydrogenase is present in the cytosol of the host cells.

In some embodiments, the host cells or fermenting organisms comprising a heterologous polynucleotide encoding a xylitol dehydrogenase (XDH) have an increased level of xylitol dehydrogenase activity compared to the host cells without the heterologous polynucleotide encoding the xylitol dehydrogenase, when cultivated under the same conditions. In some embodiments, the host cells have an increased level of xylitol dehydrogenase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cells without the heterologous polynucleotide encoding the xylitol dehydrogenase, when cultivated under the same conditions.

Exemplary xylitol dehydrogenases (XDHs) that can be used with the host cells and fermenting organisms, and methods of use described herein include, but are not limited to, the

Scheffersomyces stipitis xylitol dehydrogenase of SEQ ID NO: 466, the *Trichoderma reesei* xylitol dehydrogenase (Wang et al., 1998, *Chin. J. Biotechnol.* 14, 179-185), the *Pichia stipitis* xylitol dehydrogenase (Karhumaa et al, 2007, *Microb Cell Fact.* 6, 5), as well as other yeast xylitol dehydrogenases described in the art, such as XDHs from *S. cerevisiae* (Richard et. al., 1999, *FEBS Letters* 457, 135-138), *C. didensiae*, *C. intermediae*, *C. parapsilosis*, *C. silvanoru*, *C. tropicalis*, *Kl. Marxsianus*, *P. guilliermondii*, *T. molishiama*, *Pa. tannophilus*, and *C. shehatae* (Yablochkova et al, 2003, *Microbiology* 72(4), 414-417). Additional polynucleotides encoding suitable xylitol dehydrogenases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org). In one embodiment, the xylitol dehydrogenase is a bacterial, a yeast, or a filamentous fungal xylitol dehydrogenase, e.g., obtained from any of the microorganisms described or referenced herein, as described *supra*.

The xylitol dehydrogenase (XDH) coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding xylitol dehydrogenases from strains of different genera or species, as described *supra*.

The polynucleotides encoding xylitol dehydrogenases (XDHs) may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding xylitol dehydrogenases (XDHs) are described *supra*.

In one embodiment, the xylitol dehydrogenase (XDH) has a mature polypeptide sequence of at least 60%, e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any xylitol dehydrogenase described or referenced herein (e.g., the *Scheffersomyces stipitis* xylitol dehydrogenase of SEQ ID NO: 466). In one embodiment, the xylitol dehydrogenase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any xylitol dehydrogenase described or referenced herein (e.g., the *Scheffersomyces stipitis* xylitol dehydrogenase of SEQ ID NO: 466). In one embodiment, the xylitol dehydrogenase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any xylitol dehydrogenase described or referenced herein (e.g., the *Scheffersomyces stipitis* xylitol dehydrogenase of SEQ ID NO: 466), allelic variant, or a fragment thereof having xylitol dehydrogenase activity. In one embodiment, the xylitol dehydrogenase has

CA 03158982 2022-5-19

an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) amino acids. In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the xylitol dehydrogenase (XDH) has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the xylitol dehydrogenase activity of any xylitol dehydrogenase described or referenced herein (e.g., the *Scheffersomyces stipitis* xylitol dehydrogenase of SEQ ID NO: 466) under the same conditions.

In one embodiment, the xylitol dehydrogenase (XDH) coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any xylitol dehydrogenase described or referenced herein (e.g., the *Scheffersomyces stipitis* xylitol dehydrogenase of SEQ ID NO: 466). In one embodiment, the xylitol dehydrogenase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any xylitol dehydrogenase described or referenced herein (e.g., the *Scheffersomyces stipitis* xylitol dehydrogenase of SEQ ID NO: 466).

In one embodiment, the heterologous polynucleotide encoding the xylitol dehydrogenase (XDH) comprises the coding sequence of any xylitol dehydrogenase described or referenced herein (e.g., the *Scheffersomyces stipitis* xylitol dehydrogenase of SEQ ID NO: 466). In one embodiment, the heterologous polynucleotide encoding the xylitol dehydrogenase comprises a subsequence of the coding sequence from any xylitol dehydrogenase described or referenced herein, wherein the subsequence encodes a polypeptide having xylitol dehydrogenase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The xylitol dehydrogenases (XDHs) can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

In one embodiment, the host cell or fermenting organism (e.g., yeast cell) further comprises a heterologous polynucleotide encoding a xylulokinase (XK). A xylulokinase, as used herein, provides enzymatic activity for converting D-xylulose to xylulose 5-phosphate. The xylulokinase may be any xylulokinase that is suitable for the host cells and the methods described

CA 03158982 2022-5-19

herein, such as a naturally occurring xylulokinase or a variant thereof that retains xylulokinase activity. In one embodiment, the xylulokinase is present in the cytosol of the host cells.

In some embodiments, the host cells or fermenting organisms comprising a heterologous polynucleotide encoding a xylulokinase (XK) have an increased level of xylulokinase activity compared to the host cells without the heterologous polynucleotide encoding the xylulokinase, when cultivated under the same conditions. In some embodiments, the host cells have an increased level of xylulokinase activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cells without the heterologous polynucleotide encoding the xylulokinase, when cultivated under the same conditions.

Exemplary xylulokinases (XKs) that can be used with the host cells and fermenting organisms, and methods of use described herein include, but are not limited to, the *Saccharomyces cerevisiae* xylulokinase of SEQ ID NO: 75, the *Scheffersomyces stipitis* xylulokinase of SEQ ID NO: 467 and the *Aspergillus niger* xylulokinase of SEQ ID NO: 468. Additional xylulokinases are known in the art. Additional polynucleotides encoding suitable xylulokinases may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org). In one embodiment, the xylulokinases is a bacterial, a yeast, or a filamentous fungal xylulokinase, e.g., obtained from any of the microorganisms described or referenced herein, as described *supra*.

The xylulokinase (XK) coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding xylulokinases from strains of different genera or species, as described *supra*.

The polynucleotides encoding xylulokinases (XK) may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding xylulokinases (XKs) are described *supra*.

In one embodiment, the xylulokinase (XK) has a mature polypeptide sequence of at least 60%, e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any xylulokinase described or referenced herein (e.g., the xylulokinase of SEQ ID NO: 75, 467 or 468). In one embodiment, the xylulokinase has a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more

CA 03158982 2022-5-19

than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any xylulokinase described or referenced herein (e.g., the xylulokinase of SEQ ID NO: 75, 467 or 468). In one embodiment, the xylulokinase has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any xylulokinase described or referenced herein (e.g., the xylulokinase of SEQ ID NO: 75, 467 or 468), allelic variant, or a fragment thereof having xylulokinase activity. In one embodiment, the xylulokinase has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) amino acids. In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the xylulokinase (XK) has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the xylulokinase activity of any xylulokinase described or referenced herein (e.g., the xylulokinase of SEQ ID NO: 75, 467 or 468) under the same conditions.

In one embodiment, the xylulokinase (XK) coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any xylulokinase described or referenced herein (e.g., the xylulokinase of SEQ ID NO: 75, 467 or 468). In one embodiment, the xylulokinase coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any xylulokinase described or referenced herein (e.g., the xylulokinase of SEQ ID NO: 75, 467 or 468).

In one embodiment, the heterologous polynucleotide encoding the xylulokinase (XK) comprises the coding sequence of any xylulokinase described or referenced herein (e.g., the xylulokinase of SEQ ID NO: 75, 467 or 468). In one embodiment, the heterologous polynucleotide encoding the xylulokinase comprises a subsequence of the coding sequence from any xylulokinase described or referenced herein, wherein the subsequence encodes a polypeptide having xylulokinase activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

CA 03158982 2022-5-19

The xylulokinases (XKs) can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

In one aspect, the recombinant cells described herein (e.g., a cell comprising a heterologous polynucleotide encoding a sugar transporter) have improved anaerobic growth on a pentose (e.g., xylose and/or arabinose). In one embodiment, the recombinant cell is capable of higher anaerobic growth rate on a pentose (e.g., xylose and/or arabinose) compared to the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2).

In one aspect, the recombinant cells described herein (e.g., a cell comprising a heterologous polynucleotide encoding a sugar transporter) have improved rate of pentose consumption (e.g., xylose and/or arabinose). In one embodiment, the recombinant cell is capable of higher rate of pentose consumption (e.g., xylose and/or arabinose) compared to the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2). In one embodiment, the rate of pentose consumption (e.g., xylose and/or arabinose) is at least 5%, e.g., at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 75% or 90% higher compared to the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2).

In one aspect, the recombinant cells described herein (e.g., a cell comprising a heterologous polynucleotide encoding a sugar transporter described herein) have higher pentose (e.g., xylose and/or arabinose) consumption. In one embodiment, the recombinant cell is capable of higher pentose (e.g., xylose and/or arabinose) consumption compared to the same cell without the heterologous polynucleotide encoding a hexose transporter at about or after 120 hours fermentation (e.g., under conditions described in Example 2). In one embodiment, the recombinant cell is capable of consuming more than 65%, e.g., at least 70%, 75%, 80%, 85%, 90%, 95% of pentose (e.g., xylose and/or arabinose) in the medium at about or after 120 hours fermentation (e.g., under conditions described in Example 2).

Methods using a Starch-Containing Material

In some embodiments, the methods described herein produce a fermentation product from a starch-containing material. Starch-containing material is well-known in the art, containing two types of homopolysaccharides (amylose and amylopectin) and is linked by alpha-(1-4)-D-glycosidic bonds. Any suitable starch-containing starting material may be used. The starting material is generally selected based on the desired fermentation product, such as ethanol. Examples of starch-containing starting materials include cereal, tubers or grains. Specifically, the

CA 03158982 2022-5-19

starch-containing material may be corn, wheat, barley, rye, milo, sago, cassava, tapioca, sorghum, oat, rice, peas, beans, or sweet potatoes, or mixtures thereof. Contemplated are also waxy and non-waxy types of corn and barley.

In one embodiment, the starch-containing starting material is corn. In one embodiment, the starch-containing starting material is wheat. In one embodiment, the starch-containing starting material is barley. In one embodiment, the starch-containing starting material is rye. In one embodiment, the starch-containing starting material is milo. In one embodiment, the starch-containing starting material is sago. In one embodiment, the starch-containing starting material is cassava. In one embodiment, the starch-containing starting material is tapioca. In one embodiment, the starch-containing starting material is sorghum. In one embodiment, the starch-containing starting material is rice. In one embodiment, the starch-containing starting material is peas. In one embodiment, the starch-containing starting material is beans. In one embodiment, the starch-containing starting material is sweet potatoes. In one embodiment, the starch-containing starting material is oats.

The methods using a starch-containing material may include a conventional process (e.g., including a liquefaction step described in more detail below) or a raw starch hydrolysis process. In some embodiments using a starch-containing material, saccharification of the starch-containing material is at a temperature above the initial gelatinization temperature. In some embodiments using a starch-containing material, saccharification of the starch-containing material is at a temperature below the initial gelatinization temperature.

Liquefaction

In embodiments using a starch-containing material, the methods may further comprise a liquefaction step carried out by subjecting the starch-containing material at a temperature above the initial gelatinization temperature to an alpha-amylase and optionally a protease and/or a glucoamylase. Other enzymes such as a pullulanase and phytase may also be present and/or added in liquefaction. In some embodiments, the liquefaction step is carried out prior to steps a) and b) of the described methods.

Liquefaction step may be carried out for 0.5-5 hours, such as 1-3 hours, such as typically about 2 hours.

The term "initial gelatinization temperature" means the lowest temperature at which gelatinization of the starch-containing material commences. In general, starch heated in water begins to gelatinize between about 50°C and 75°C; the exact temperature of gelatinization depends on the specific starch and can readily be determined by the skilled artisan. Thus, the

initial gelatinization temperature may vary according to the plant species, to the particular variety of the plant species as well as with the growth conditions. The initial gelatinization temperature of a given starch-containing material may be determined as the temperature at which birefringence is lost in 5% of the starch granules using the method described by Gorinstein and Lii, 1992, *Starch/Stärke* 44(12): 461-466.

Liquefaction is typically carried out at a temperature in the range from 70-100ºC. In one embodiment, the temperature in liquefaction is between 75-95°C, such as between 75-90°C, between 80-90°C, or between 82-88°C, such as about 85ºC.

A jet-cooking step may be carried out prior to liquefaction in step, for example, at a temperature between 110-145°C, 120-140ºC, 125-135°C, or about 130ºC for about 1-15 minutes, for about 3-10 minutes, or about 5 minutes.

The pH during liquefaction may be between 4 and 7, such as pH 4.5-6.5, pH 5.0-6.5, pH 5.0-6.0, pH 5.2-6.2, or about 5.2, about 5.4, about 5.6, or about 5.8.

In one embodiment, the process further comprises, prior to liquefaction, the steps of:

i) reducing the particle size of the starch-containing material, preferably by dry milling;

ii) forming a slurry comprising the starch-containing material and water.

The starch-containing starting material, such as whole grains, may be reduced in particle size, e.g., by milling, in order to open up the structure, to increase surface area, and allowing for further processing. Generally, there are two types of processes: wet and dry milling. In dry milling whole kernels are milled and used. Wet milling gives a good separation of germ and meal (starch granules and protein). Wet milling is often applied at locations where the starch hydrolysate is used in production of, e.g., syrups. Both dry milling and wet milling are well known in the art of starch processing. In one embodiment the starch-containing material is subjected to dry milling. In one embodiment, the particle size is reduced to between 0.05 to 3.0 mm, e.g., 0.1-0.5 mm, or so that at least 30%, at least 50%, at least 70%, or at least 90% of the starch-containing material fit through a sieve with a 0.05 to 3.0 mm screen, e.g., 0.1-0.5 mm screen. In another embodiment, at least 50%, e.g., at least 70%, at least 80%, or at least 90% of the starch-containing material fit through a sieve with # 6 screen.

The aqueous slurry may contain from 10-55 w/w-% dry solids (DS), e.g., 25-45 w/w-% dry solids (DS), or 30-40 w/w-% dry solids (DS) of starch-containing material.

The alpha-amylase, optionally a protease, and optionally a glucoamylase may initially be added to the aqueous slurry to initiate liquefaction (thinning). In one embodiment, only a portion of the enzymes (e.g., about 1/3) is added to the aqueous slurry, while the rest of the enzymes (e.g., about 2/3) are added during liquefaction step.

CA 03158982 2022-5-19

A non-exhaustive list of alpha-amylases used in liquefaction can be found in the "Alpha-Amylases" section. Examples of suitable proteases used in liquefaction include any protease described *supra* in the "Proteases" section. Examples of suitable glucoamylases used in liquefaction include any glucoamylase found in the "Glucoamylases" section.

Saccharification and Fermentation of Starch-containing material

In embodiments using a starch-containing material, a glucoamylase may be present and/or added in saccharification step a) and/or fermentation step b) or simultaneous saccharification and fermentation (SSF). The glucoamylase of the saccharification step a) and/or fermentation step b) or simultaneous saccharification and fermentation (SSF) is typically different from the glucoamylase optionally added to any liquefaction step described *supra*. In one embodiment, the glucoamylase is present and/or added together with a fungal alpha-amylase.

In some embodiments, the host cell or fermenting organism comprises a heterologous polynucleotide encoding a glucoamylase, for example, as described in WO2017/087330, the content of which is hereby incorporated by reference.

Examples of glucoamylases can be found in the "Glucoamylases" section.

When doing sequential saccharification and fermentation, saccharification step a) may be carried out under conditions well-known in the art. For instance, saccharification step a) may last up to from about 24 to about 72 hours. In one embodiment, pre-saccharification is done. Pre-saccharification is typically done for 40-90 minutes at a temperature between 30-65°C, typically about 60°C. Pre-saccharification is, in one embodiment, followed by saccharification during fermentation in simultaneous saccharification and fermentation (SSF). Saccharification is typically carried out at temperatures from 20-75°C, preferably from 40-70°C, typically about 60°C, and typically at a pH between 4 and 5, such as about pH 4.5.

Fermentation is carried out in a fermentation medium, as known in the art and, e.g., as described herein. The fermentation medium includes the fermentation substrate, that is, the carbohydrate source that is metabolized by the fermenting organism. With the processes described herein, the fermentation medium may comprise nutrients and growth stimulator(s) for the fermenting organism(s). Nutrient and growth stimulators are widely used in the art of fermentation and include nitrogen sources, such as ammonia; urea, vitamins and minerals, or combinations thereof.

Generally, fermenting organisms such as yeast, including *Saccharomyces cerevisiae* yeast, require an adequate source of nitrogen for propagation and fermentation. Many sources of supplemental nitrogen, if necessary, can be used and such sources of nitrogen are well known

CA 03158982 2022-5-19

in the art. The nitrogen source may be organic, such as urea, DDGs, wet cake or corn mash, or inorganic, such as ammonia or ammonium hydroxide. In one embodiment, the nitrogen source is urea.

Fermentation can be carried out under low nitrogen conditions, e.g., when using a protease-expressing yeast. In some embodiments, the fermentation step is conducted with less than 1000 ppm supplemental nitrogen (e.g., urea or ammonium hydroxide), such as less than 750 ppm, less than 500 ppm, less than 400 ppm, less than 300 ppm, less than 250 ppm, less than 200 ppm, less than 150 ppm, less than 100 ppm, less than 75 ppm, less than 50 ppm, less than 25 ppm, or less than 10 ppm, supplemental nitrogen. In some embodiments, the fermentation step is conducted with no supplemental nitrogen.

Simultaneous saccharification and fermentation ("SSF") is widely used in industrial scale fermentation product production processes, especially ethanol production processes. When doing SSF the saccharification step a) and the fermentation step b) are carried out simultaneously. There is no holding stage for the saccharification, meaning that a fermenting organism, such as yeast, and enzyme(s), may be added together. However, it is also contemplated to add the fermenting organism and enzyme(s) separately. SSF is typically carried out at a temperature from 25°C to 40°C, such as from 28°C to 35°C, such as from 30°C to 34°C, or about 32°C. In one embodiment, fermentation is ongoing for 6 to 120 hours, in particular 24 to 96 hours. In one embodiment, the pH is between 4-5.

In one embodiment, a cellulolytic enzyme composition is present and/or added in saccharification, fermentation or simultaneous saccharification and fermentation (SSF). Examples of such cellulolytic enzyme compositions can be found in the "Cellulolytic Enzymes and Compositions" section. The cellulolytic enzyme composition may be present and/or added together with a glucoamylase, such as one disclosed in the "Glucoamylases" section.

Methods using a Cellulosic-Containing Material

In some embodiments, the methods described herein produce a fermentation product from a cellulosic-containing material. The predominant polysaccharide in the primary cell wall of biomass is cellulose, the second most abundant is hemicellulose, and the third is pectin. The secondary cell wall, produced after the cell has stopped growing, also contains polysaccharides and is strengthened by polymeric lignin covalently cross-linked to hemicellulose. Cellulose is a homopolymer of anhydrocellobiose and thus a linear beta-(1-4)-D-glucan, while hemicelluloses include a variety of compounds, such as xylans, xyloglucans, arabinoxylans, and mannans in complex branched structures with a spectrum of substituents. Although generally polymorphous,

CA 03158982 2022-5-19

cellulose is found in plant tissue primarily as an insoluble crystalline matrix of parallel glucan chains. Hemicelluloses usually hydrogen bond to cellulose, as well as to other hemicelluloses, which help stabilize the cell wall matrix.

Cellulose is generally found, for example, in the stems, leaves, hulls, husks, and cobs of plants or leaves, branches, and wood of trees. The cellulosic-containing material can be, but is not limited to, agricultural residue, herbaceous material (including energy crops), municipal solid waste, pulp and paper mill residue, waste paper, and wood (including forestry residue) (see, for example, Wiselogel *et al.*, 1995, in Handbook on Bioethanol (Charles E. Wyman, editor), pp. 105-118, Taylor & Francis, Washington D.C.; Wyman, 1994, *Bioresource Technology* 50: 3-16; Lynd, 1990, *Applied Biochemistry and Biotechnology* 24/25: 695-719; Mosier *et al.*, 1999, Recent Progress in Bioconversion of Lignocellulosics, in *Advances in Biochemical Engineering/Biotechnology*, T. Scheper, managing editor, Volume 65, pp. 23-40, Springer-Verlag, New York). It is understood herein that the cellulose may be in the form of lignocellulose, a plant cell wall material containing lignin, cellulose, and hemicellulose in a mixed matrix. In one embodiment, the cellulosic-containing material is any biomass material. In another embodiment, the cellulosic-containing material is lignocellulose, which comprises cellulose, hemicelluloses, and lignin.

In one embodiment, the cellulosic-containing material is agricultural residue, herbaceous material (including energy crops), municipal solid waste, pulp and paper mill residue, waste paper, or wood (including forestry residue).

In another embodiment, the cellulosic-containing material is arundo, bagasse, bamboo, corn cob, corn fiber, corn stover, miscanthus, rice straw, switchgrass, or wheat straw.

In another embodiment, the cellulosic-containing material is aspen, eucalyptus, fir, pine, poplar, spruce, or willow.

In another embodiment, the cellulosic-containing material is algal cellulose, bacterial cellulose, cotton linter, filter paper, microcrystalline cellulose (e.g., AVICEL®), or phosphoric-acid treated cellulose.

In another embodiment, the cellulosic-containing material is an aquatic biomass. As used herein the term "aquatic biomass" means biomass produced in an aquatic environment by a photosynthesis process. The aquatic biomass can be algae, emergent plants, floating-leaf plants, or submerged plants.

The cellulosic-containing material may be used as is or may be subjected to pretreatment, using conventional methods known in the art, as described herein. In a preferred embodiment, the cellulosic-containing material is pretreated.

CA 03158982 2022-5-19

The methods of using cellulosic-containing material can be accomplished using methods conventional in the art. Moreover, the methods of can be implemented using any conventional biomass processing apparatus configured to carry out the processes.

Cellulosic Pretreatment

In one embodiment the cellulosic-containing material is pretreated before saccharification.

In practicing the processes described herein, any pretreatment process known in the art can be used to disrupt plant cell wall components of the cellulosic-containing material (Chandra *et al.*, 2007, *Adv. Biochem. Engin./Biotechnol.* 108: 67-93; Galbe and Zacchi, 2007, *Adv. Biochem. Engin./Biotechnol.* 108: 41-65; Hendriks and Zeeman, 2009, *Bioresource Technology* 100: 10-18; Mosier *et al.*, 2005, *Bioresource Technology* 96: 673-686; Taherzadeh and Karimi, 2008, *Int. J. Mol. Sci.* 9: 1621-1651; Yang and Wyman, 2008, *Biofuels Bioproducts and Biorefining-Biofpr.* 2: 26-40).

The cellulosic-containing material can also be subjected to particle size reduction, sieving, pre-soaking, wetting, washing, and/or conditioning prior to pretreatment using methods known in the art.

Conventional pretreatments include, but are not limited to, steam pretreatment (with or without explosion), dilute acid pretreatment, hot water pretreatment, alkaline pretreatment, lime pretreatment, wet oxidation, wet explosion, ammonia fiber explosion, organosolv pretreatment, and biological pretreatment. Additional pretreatments include ammonia percolation, ultrasound, electroporation, microwave, supercritical CO_2, supercritical H_2O, ozone, ionic liquid, and gamma irradiation pretreatments.

In a one embodiment, the cellulosic-containing material is pretreated before saccharification (i.e., hydrolysis) and/or fermentation. Pretreatment is preferably performed prior to the hydrolysis. Alternatively, the pretreatment can be carried out simultaneously with enzyme hydrolysis to release fermentable sugars, such as glucose, xylose, and/or cellobiose. In most cases the pretreatment step itself results in some conversion of biomass to fermentable sugars (even in absence of enzymes).

In one embodiment, the cellulosic-containing material is pretreated with steam. In steam pretreatment, the cellulosic-containing material is heated to disrupt the plant cell wall components, including lignin, hemicellulose, and cellulose to make the cellulose and other fractions, e.g., hemicellulose, accessible to enzymes. The cellulosic-containing material is passed to or through a reaction vessel where steam is injected to increase the temperature to the required temperature and pressure and is retained therein for the desired reaction time. Steam pretreatment is

CA 03158982 2022-5-19

preferably performed at 140-250°C, e.g., 160-200°C or 170-190°C, where the optimal temperature range depends on optional addition of a chemical catalyst. Residence time for the steam pretreatment is preferably 1-60 minutes, e.g., 1-30 minutes, 1-20 minutes, 3-12 minutes, or 4-10 minutes, where the optimal residence time depends on the temperature and optional addition of a chemical catalyst. Steam pretreatment allows for relatively high solids loadings, so that the cellulosic-containing material is generally only moist during the pretreatment. The steam pretreatment is often combined with an explosive discharge of the material after the pretreatment, which is known as steam explosion, that is, rapid flashing to atmospheric pressure and turbulent flow of the material to increase the accessible surface area by fragmentation (Duff and Murray, 1996, *Bioresource Technology* 855: 1-33; Galbe and Zacchi, 2002, *Appl. Microbiol. Biotechnol.* 59: 618-628; U.S. Patent Application No. 2002/0164730). During steam pretreatment, hemicellulose acetyl groups are cleaved and the resulting acid autocatalyzes partial hydrolysis of the hemicellulose to monosaccharides and oligosaccharides. Lignin is removed to only a limited extent.

In one embodiment, the cellulosic-containing material is subjected to a chemical pretreatment. The term "chemical treatment" refers to any chemical pretreatment that promotes the separation and/or release of cellulose, hemicellulose, and/or lignin. Such a pretreatment can convert crystalline cellulose to amorphous cellulose. Examples of suitable chemical pretreatment processes include, for example, dilute acid pretreatment, lime pretreatment, wet oxidation, ammonia fiber/freeze expansion (AFEX), ammonia percolation (APR), ionic liquid, and organosolv pretreatments.

A chemical catalyst such as H_2SO_4 or SO_2 (typically 0.3 to 5% w/w) is sometimes added prior to steam pretreatment, which decreases the time and temperature, increases the recovery, and improves enzymatic hydrolysis (Ballesteros *et al.*, 2006, *Appl. Biochem. Biotechnol.* 129-132: 496-508; Varga *et al.*, 2004, *Appl. Biochem. Biotechnol.* 113-116: 509-523; Sassner *et al.*, 2006, *Enzyme Microb. Technol.* 39: 756-762). In dilute acid pretreatment, the cellulosic-containing material is mixed with dilute acid, typically H_2SO_4, and water to form a slurry, heated by steam to the desired temperature, and after a residence time flashed to atmospheric pressure. The dilute acid pretreatment can be performed with a number of reactor designs, e.g., plug-flow reactors, counter-current reactors, or continuous counter-current shrinking bed reactors (Duff and Murray, 1996, *Bioresource Technology* 855: 1-33; Schell *et al.*, 2004, *Bioresource Technology* 91: 179-188; Lee *et al.*, 1999, *Adv. Biochem. Eng. Biotechnol.* 65: 93-115). In a specific embodiment the dilute acid pretreatment of cellulosic-containing material is carried out using 4% w/w sulfuric acid at 180°C for 5 minutes.

CA 03158982 2022-5-19

Several methods of pretreatment under alkaline conditions can also be used. These alkaline pretreatments include, but are not limited to, sodium hydroxide, lime, wet oxidation, ammonia percolation (APR), and ammonia fiber/freeze expansion (AFEX) pretreatment. Lime pretreatment is performed with calcium oxide or calcium hydroxide at temperatures of 85-150°C and residence times from 1 hour to several days (Wyman *et al.*, 2005, *Bioresource Technology* 96: 1959-1966; Mosier *et al.*, 2005, *Bioresource Technology* 96: 673-686). WO 2006/110891, WO 2006/110899, WO 2006/110900, and WO 2006/110901 disclose pretreatment methods using ammonia.

Wet oxidation is a thermal pretreatment performed typically at 180-200°C for 5-15 minutes with addition of an oxidative agent such as hydrogen peroxide or over-pressure of oxygen (Schmidt and Thomsen, 1998, *Bioresource Technology* 64: 139-151; Palonen *et al.*, 2004, *Appl. Biochem. Biotechnol.* 117: 1-17; Varga *et al.*, 2004, *Biotechnol. Bioeng.* 88: 567-574; Martin *et al.*, 2006, *J. Chem. Technol. Biotechnol.* 81: 1669-1677). The pretreatment is performed preferably at 1-40% dry matter, e.g., 2-30% dry matter or 5-20% dry matter, and often the initial pH is increased by the addition of alkali such as sodium carbonate.

A modification of the wet oxidation pretreatment method, known as wet explosion (combination of wet oxidation and steam explosion) can handle dry matter up to 30%. In wet explosion, the oxidizing agent is introduced during pretreatment after a certain residence time. The pretreatment is then ended by flashing to atmospheric pressure (WO 2006/032282).

Ammonia fiber expansion (AFEX) involves treating the cellulosic-containing material with liquid or gaseous ammonia at moderate temperatures such as 90-150°C and high pressure such as 17-20 bar for 5-10 minutes, where the dry matter content can be as high as 60% (Gollapalli *et al.*, 2002, *Appl. Biochem. Biotechnol.* 98: 23-35; Chundawat *et al.*, 2007, *Biotechnol. Bioeng.* 96: 219-231; Alizadeh *et al.*, 2005, *Appl. Biochem. Biotechnol.* 121: 1133-1141; Teymouri *et al.*, 2005, *Bioresource Technology* 96: 2014-2018). During AFEX pretreatment cellulose and hemicelluloses remain relatively intact. Lignin-carbohydrate complexes are cleaved.

Organosolv pretreatment delignifies the cellulosic-containing material by extraction using aqueous ethanol (40-60% ethanol) at 160-200°C for 30-60 minutes (Pan *et al.*, 2005, *Biotechnol. Bioeng.* 90: 473-481; Pan *et al.*, 2006, *Biotechnol. Bioeng.* 94: 851-861; Kurabi *et al.*, 2005, *Appl. Biochem. Biotechnol.* 121: 219-230). Sulphuric acid is usually added as a catalyst. In organosolv pretreatment, the majority of hemicellulose and lignin is removed.

Other examples of suitable pretreatment methods are described by Schell *et al.*, 2003, *Appl. Biochem. Biotechnol.* 105-108: 69-85, and Mosier *et al.*, 2005, *Bioresource Technology* 96: 673-686, and U.S. Published Application 2002/0164730.

CA 03158982 2022-5-19

In one embodiment, the chemical pretreatment is carried out as a dilute acid treatment, and more preferably as a continuous dilute acid treatment. The acid is typically sulfuric acid, but other acids can also be used, such as acetic acid, citric acid, nitric acid, phosphoric acid, tartaric acid, succinic acid, hydrogen chloride, or mixtures thereof. Mild acid treatment is conducted in the pH range of preferably 1-5, e.g., 1-4 or 1-2.5. In one embodiment, the acid concentration is in the range from preferably 0.01 to 10 wt. % acid, e.g., 0.05 to 5 wt. % acid or 0.1 to 2 wt. % acid. The acid is contacted with the cellulosic-containing material and held at a temperature in the range of preferably 140-200°C, e.g., 165-190°C, for periods ranging from 1 to 60 minutes.

In another embodiment, pretreatment takes place in an aqueous slurry. In preferred embodiments, the cellulosic-containing material is present during pretreatment in amounts preferably between 10-80 wt. %, e.g., 20-70 wt. % or 30-60 wt. %, such as around 40 wt. %. The pretreated cellulosic-containing material can be unwashed or washed using any method known in the art, e.g., washed with water.

In one embodiment, the cellulosic-containing material is subjected to mechanical or physical pretreatment. The term "mechanical pretreatment" or "physical pretreatment" refers to any pretreatment that promotes size reduction of particles. For example, such pretreatment can involve various types of grinding or milling (e.g., dry milling, wet milling, or vibratory ball milling).

The cellulosic-containing material can be pretreated both physically (mechanically) and chemically. Mechanical or physical pretreatment can be coupled with steaming/steam explosion, hydrothermolysis, dilute or mild acid treatment, high temperature, high pressure treatment, irradiation (e.g., microwave irradiation), or combinations thereof. In one embodiment, high pressure means pressure in the range of preferably about 100 to about 400 psi, e.g., about 150 to about 250 psi. In another embodiment, high temperature means temperature in the range of about 100 to about 300°C, e.g., about 140 to about 200°C. In a preferred embodiment, mechanical or physical pretreatment is performed in a batch-process using a steam gun hydrolyzer system that uses high pressure and high temperature as defined above, e.g., a Sunds Hydrolyzer available from Sunds Defibrator AB, Sweden. The physical and chemical pretreatments can be carried out sequentially or simultaneously, as desired.

Accordingly, in one embodiment, the cellulosic-containing material is subjected to physical (mechanical) or chemical pretreatment, or any combination thereof, to promote the separation and/or release of cellulose, hemicellulose, and/or lignin.

In one embodiment, the cellulosic-containing material is subjected to a biological pretreatment. The term "biological pretreatment" refers to any biological pretreatment that promotes the separation and/or release of cellulose, hemicellulose, and/or lignin from the

CA 03158982 2022-5-19

cellulosic-containing material. Biological pretreatment techniques can involve applying lignin-solubilizing microorganisms and/or enzymes (see, for example, Hsu, T.-A., 1996, Pretreatment of biomass, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212; Ghosh and Singh, 1993, *Adv. Appl. Microbiol.* 39: 295-333; McMillan, J. D., 1994, Pretreating lignocellulosic biomass: a review, in *Enzymatic Conversion of Biomass for Fuels Production*, Himmel, M. E., Baker, J. O., and Overend, R. P., eds., ACS Symposium Series 566, American Chemical Society, Washington, DC, chapter 15; Gong, C. S., Cao, N. J., Du, J., and Tsao, G. T., 1999, Ethanol production from renewable resources, in *Advances in Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heidelberg, Germany, 65: 207-241; Olsson and Hahn-Hagerdal, 1996, *Enz. Microb. Tech.* 18: 312-331; and Vallander and Eriksson, 1990, *Adv. Biochem. Eng./Biotechnol.* 42: 63-95).

Saccharification and Fermentation of Cellulosic-containing material

Saccharification (i.e., hydrolysis) and fermentation, separate or simultaneous, include, but are not limited to, separate hydrolysis and fermentation (SHF); simultaneous saccharification and fermentation (SSF); simultaneous saccharification and co-fermentation (SSCF); hybrid hydrolysis and fermentation (HHF); separate hydrolysis and co-fermentation (SHCF); hybrid hydrolysis and co-fermentation (HHCF).

SHF uses separate process steps to first enzymatically hydrolyze the cellulosic-containing material to fermentable sugars, e.g., glucose, cellobiose, and pentose monomers, and then ferment the fermentable sugars to ethanol. In SSF, the enzymatic hydrolysis of the cellulosic-containing material and the fermentation of sugars to ethanol are combined in one step (Philippidis, G. P., 1996, Cellulose bioconversion technology, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212). SSCF involves the co-fermentation of multiple sugars (Sheehan and Himmel, 1999, *Biotechnol. Prog.* 15: 817-827). HHF involves a separate hydrolysis step, and in addition a simultaneous saccharification and hydrolysis step, which can be carried out in the same reactor. The steps in an HHF process can be carried out at different temperatures, *i.e.*, high temperature enzymatic saccharification followed by SSF at a lower temperature that the fermentation organismcan tolerate. It is understood herein that any method known in the art comprising pretreatment, enzymatic hydrolysis (saccharification), fermentation, or a combination thereof, can be used in the practicing the processes described herein.

A conventional apparatus can include a fed-batch stirred reactor, a batch stirred reactor, a continuous flow stirred reactor with ultrafiltration, and/or a continuous plug-flow column reactor

CA 03158982 2022-5-19

(de Castilhos Corazza *et al.*, 2003, *Acta Scientiarum. Technology* 25: 33-38; Gusakov and Sinitsyn, 1985, *Enz. Microb. Technol.* 7: 346-352), an attrition reactor (Ryu and Lee, 1983, *Biotechnol. Bioeng.* 25: 53-65). Additional reactor types include fluidized bed, upflow blanket, immobilized, and extruder type reactors for hydrolysis and/or fermentation.

In the saccharification step (i.e., hydrolysis step), the cellulosic and/or starch-containing material, e.g., pretreated, is hydrolyzed to break down cellulose, hemicellulose, and/or starch to fermentable sugars, such as glucose, cellobiose, xylose, xylulose, arabinose, mannose, galactose, and/or soluble oligosaccharides. The hydrolysis is performed enzymatically e.g., by a cellulolytic enzyme composition. The enzymes of the compositions can be added simultaneously or sequentially.

Enzymatic hydrolysis may be carried out in a suitable aqueous environment under conditions that can be readily determined by one skilled in the art. In one embodiment, hydrolysis is performed under conditions suitable for the activity of the enzymes(s), *i.e.*, optimal for the enzyme(s). The hydrolysis can be carried out as a fed batch or continuous process where the cellulosic and/or starch-containing material is fed gradually to, for example, an enzyme containing hydrolysis solution.

The saccharification is generally performed in stirred-tank reactors or fermentors under controlled pH, temperature, and mixing conditions. Suitable process time, temperature and pH conditions can readily be determined by one skilled in the art. For example, the saccharification can last up to 200 hours, but is typically performed for preferably about 12 to about 120 hours, e.g., about 16 to about 72 hours or about 24 to about 48 hours. The temperature is in the range of preferably about 25°C to about 70°C, e.g., about 30°C to about 65°C, about 40°C to about 60°C, or about 50°C to about 55°C. The pH is in the range of preferably about 3 to about 8, e.g., about 3.5 to about 7, about 4 to about 6, or about 4.5 to about 5.5. The dry solids content is in the range of preferably about 5 to about 50 wt. %, e.g., about 10 to about 40 wt. % or about 20 to about 30 wt. %.

Saccharification in may be carried out using a cellulolytic enzyme composition. Such enzyme compositions are described below in the "Cellulolytic Enzyme Composition'-section below. The cellulolytic enzyme compositions can comprise any protein useful in degrading the cellulosic-containing material. In one embodiment, the cellulolytic enzyme composition comprises or further comprises one or more (e.g., several) proteins selected from the group consisting of a cellulase, an AA9 (GH61) polypeptide, a hemicellulase, an esterase, an expansin, a ligninolytic enzyme, an oxidoreductase, a pectinase, a protease, and a swollenin.

CA 03158982 2022-5-19

In another embodiment, the cellulase is preferably one or more (e.g., several) enzymes selected from the group consisting of an endoglucanase, a cellobiohydrolase, and a beta-glucosidase.

In another embodiment, the hemicellulase is preferably one or more (e.g., several) enzymes selected from the group consisting of an acetylmannan esterase, an acetylxylan esterase, an arabinanase, an arabinofuranosidase, a coumaric acid esterase, a feruloyl esterase, a galactosidase, a glucuronidase, a glucuronoyl esterase, a mannanase, a mannosidase, a xylanase, and a xylosidase. In another embodiment, the oxidoreductase is one or more (e.g., several) enzymes selected from the group consisting of a catalase, a laccase, and a peroxidase. The enzymes or enzyme compositions used in a processes of the present invention may be in any form suitable for use, such as, for example, a fermentation broth formulation or a cell composition, a cell lysate with or without cellular debris, a semi-purified or purified enzyme preparation, or a host cell as a source of the enzymes. The enzyme composition may be a dry powder or granulate, a non-dusting granulate, a liquid, a stabilized liquid, or a stabilized protected enzyme. Liquid enzyme preparations may, for instance, be stabilized by adding stabilizers such as a sugar, a sugar alcohol or another polyol, and/or lactic acid or another organic acid according to established processes.

In one embodiment, an effective amount of cellulolytic or hemicellulolytic enzyme composition to the cellulosic-containing material is about 0.5 to about 50 mg, e.g., about 0.5 to about 40 mg, about 0.5 to about 25 mg, about 0.75 to about 20 mg, about 0.75 to about 15 mg, about 0.5 to about 10 mg, or about 2.5 to about 10 mg per g of the cellulosic-containing material.

In one embodiment, such a compound is added at a molar ratio of the compound to glucosyl units of cellulose of about 10^{-6} to about 10, e.g., about 10^{-6} to about 7.5, about 10^{-6} to about 5, about 10^{-6} to about 2.5, about 10^{-6} to about 1, about 10^{-5} to about 1, about 10^{-5} to about 10^{-1}, about 10^{-4} to about 10^{-1}, about 10^{-3} to about 10^{-1}, or about 10^{-3} to about 10^{-2}. In another embodiment, an effective amount of such a compound is about 0.1 µM to about 1 M, e.g., about 0.5 µM to about 0.75 M, about 0.75 µM to about 0.5 M, about 1 µM to about 0.25 M, about 1 µM to about 0.1 M, about 5 µM to about 50 mM, about 10 µM to about 25 mM, about 50 µM to about 25 mM, about 10 µM to about 10 mM, about 5 µM to about 5 mM, or about 0.1 mM to about 1 mM.

The term "liquor" means the solution phase, either aqueous, organic, or a combination thereof, arising from treatment of a lignocellulose and/or hemicellulose material in a slurry, or monosaccharides thereof, e.g., xylose, arabinose, mannose, etc. under conditions as described in WO 2012/021401, and the soluble contents thereof. A liquor for cellulolytic enhancement of an

CA 03158982 2022-5-19

AA9 polypeptide (GH61 polypeptide) can be produced by treating a lignocellulose or hemicellulose material (or feedstock) by applying heat and/or pressure, optionally in the presence of a catalyst, e.g., acid, optionally in the presence of an organic solvent, and optionally in combination with physical disruption of the material, and then separating the solution from the residual solids. Such conditions determine the degree of cellulolytic enhancement obtainable through the combination of liquor and an AA9 polypeptide during hydrolysis of a cellulosic substrate by a cellulolytic enzyme preparation. The liquor can be separated from the treated material using a method standard in the art, such as filtration, sedimentation, or centrifugation.

In one embodiment, an effective amount of the liquor to cellulose is about 10^{-6} to about 10 g per g of cellulose, e.g., about 10^{-6} to about 7.5 g, about 10^{-6} to about 5 g, about 10^{-6} to about 2.5 g, about 10^{-6} to about 1 g, about 10^{-5} to about 1 g, about 10^{-5} to about 10^{-1} g, about 10^{-4} to about 10^{-1} g, about 10^{-3} to about 10^{-1} g, or about 10^{-3} to about 10^{-2} g per g of cellulose.

In the fermentation step, sugars, released from the cellulosic-containing material, e.g., as a result of the pretreatment and enzymatic hydrolysis steps, are fermented to ethanol, by a host cell or fermenting organism, such as yeast described herein. Hydrolysis (saccharification) and fermentation can be separate or simultaneous.

Any suitable hydrolyzed cellulosic-containing material can be used in the fermentation step in practicing the processes described herein. Such feedstocks include, but are not limited to carbohydrates (e.g., lignocellulose, xylans, cellulose, starch, etc.). The material is generally selected based on economics, *i.e.*, costs per equivalent sugar potential, and recalcitrance to enzymatic conversion.

Production of ethanol by a host cell or fermenting organism using cellulosic-containing material results from the metabolism of sugars (monosaccharides). The sugar composition of the hydrolyzed cellulosic-containing material and the ability of the host cell or fermenting organism to utilize the different sugars has a direct impact in process yields. Prior to Applicant's disclosure herein, strains known in the art utilize glucose efficiently but do not (or very limitedly) metabolize pentoses like xylose, a monosaccharide commonly found in hydrolyzed material.

Compositions of the fermentation media and fermentation conditions depend on the host cell or fermenting organism and can easily be determined by one skilled in the art. Typically, the fermentation takes place under conditions known to be suitable for generating the fermentation product. In some embodiments, the fermentation process is carried out under aerobic or microaerophilic (i.e., where the concentration of oxygen is less than that in air), or anaerobic conditions. In some embodiments, fermentation is conducted under anaerobic conditions (i.e., no detectable oxygen), or less than about 5, about 2.5, or about 1 mmol/L/h oxygen. In the absence

CA 03158982 2022-5-19

of oxygen, the NADH produced in glycolysis cannot be oxidized by oxidative phosphorylation. Under anaerobic conditions, pyruvate or a derivative thereof may be utilized by the host cell as an electron and hydrogen acceptor in order to generate NAD+.

The fermentation process is typically run at a temperature that is optimal for the recombinant fungal cell. For example, in some embodiments, the fermentation process is performed at a temperature in the range of from about 25°C to about 42°C. Typically the process is carried out a temperature that is less than about 38°C, less than about 35°C, less than about 33°C, or less than about 38°C, but at least about 20°C, 22°C, or 25°C.

A fermentation stimulator can be used in a process described herein to further improve the fermentation, and in particular, the performance of the host cell or fermenting organism, such as, rate enhancement and product yield (e.g., ethanol yield). A "fermentation stimulator" refers to stimulators for growth of the host cells and fermenting organisms, in particular, yeast. Preferred fermentation stimulators for growth include vitamins and minerals. Examples of vitamins include multivitamins, biotin, pantothenate, nicotinic acid, meso-inositol, thiamine, pyridoxine, para-aminobenzoic acid, folic acid, riboflavin, and Vitamins A, B, C, D, and E. See, for example, Alfenore *et al.*, Improving ethanol production and viability of *Saccharomyces cerevisiae* by a vitamin feeding strategy during fed-batch process, Springer-Verlag (2002), which is hereby incorporated by reference. Examples of minerals include minerals and mineral salts that can supply nutrients comprising P, K, Mg, S, Ca, Fe, Zn, Mn, and Cu.

Cellulolytic Enzymes and Compositions

A cellulolytic enzyme or cellulolytic enzyme composition may be present and/or added during saccharification. A cellulolytic enzyme composition is an enzyme preparation containing one or more (e.g., several) enzymes that hydrolyze cellulosic-containing material. Such enzymes include endoglucanase, cellobiohydrolase, beta-glucosidase, and/or combinations thereof.

In some embodiments, the host cell or fermenting organism comprises one or more (e.g., several) heterologous polynucleotides encoding enzymes that hydrolyze cellulosic-containing material (e.g., an endoglucanase, cellobiohydrolase, beta-glucosidase or combinations thereof). Any enzyme described or referenced herein that hydrolyzes cellulosic-containing material is contemplated for expression in the host cell or fermenting organism.

The cellulolytic enzyme may be any cellulolytic enzyme that is suitable for the host cells and/or the methods described herein (e.g., an endoglucanase, cellobiohydrolase, beta-glucosidase), such as a naturally occurring cellulolytic enzyme or a variant thereof that retains cellulolytic enzyme activity.

CA 03158982 2022-5-19

In some embodiments, the host cell or fermenting organism comprising a heterologous polynucleotide encoding a cellulolytic enzyme has an increased level of cellulolytic enzyme activity (e.g., increased endoglucanase, cellobiohydrolase, and/or beta-glucosidase) compared to the host cells without the heterologous polynucleotide encoding the cellulolytic enzyme, when cultivated under the same conditions. In some embodiments, the host cell or fermenting organism has an increased level of cellulolytic enzyme activity of at least 5%, e.g., at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, at least 100%, at least 150%, at least 200%, at least 300%, or at 500% compared to the host cell or fermenting organism without the heterologous polynucleotide encoding the cellulolytic enzyme, when cultivated under the same conditions.

Exemplary cellulolytic enzymes that can be used with the host cells and/or the methods described herein include bacterial, yeast, or filamentous fungal cellulolytic enzymes, e.g., obtained from any of the microorganisms described or referenced herein, as described *supra* under the sections related to proteases.

The cellulolytic enzyme may be of any origin. In an embodiment the cellulolytic enzyme is derived from a strain of *Trichoderma*, such as a strain of *Trichoderma reesei*; a strain of *Humicola*, such as a strain of *Humicola insolens*, and/or a strain of *Chrysosporium*, such as a strain of *Chrysosporium lucknowense*. In a preferred embodiment the cellulolytic enzyme is derived from a strain of *Trichoderma reesei*.

The cellulolytic enzyme composition may further comprise one or more of the following polypeptides, such as enzymes: AA9 polypeptide (GH61 polypeptide) having cellulolytic enhancing activity, beta-glucosidase, xylanase, beta-xylosidase, CBH I, CBH II, or a mixture of two, three, four, five or six thereof.

The further polypeptide(s) (e.g., AA9 polypeptide) and/or enzyme(s) (e.g., beta-glucosidase, xylanase, beta-xylosidase, CBH I and/or CBH II may be foreign to the cellulolytic enzyme composition producing organism (e.g., *Trichoderma reesei*).

In an embodiment the cellulolytic enzyme composition comprises an AA9 polypeptide having cellulolytic enhancing activity and a beta-glucosidase.

In another embodiment the cellulolytic enzyme composition comprises an AA9 polypeptide having cellulolytic enhancing activity, a beta-glucosidase, and a CBH I.

In another embodiment the cellulolytic enzyme composition comprises an AA9 polypeptide having cellulolytic enhancing activity, a beta-glucosidase, a CBH I and a CBH II.

Other enzymes, such as endoglucanases, may also be comprised in the cellulolytic enzyme composition.

CA 03158982 2022-5-19

As mentioned above the cellulolytic enzyme composition may comprise a number of difference polypeptides, including enzymes.

In one embodiment, the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Thermoascus aurantiacus* AA9 (GH61A) polypeptide having cellulolytic enhancing activity (e.g., WO 2005/074656), and *Aspergillus oryzae* beta-glucosidase fusion protein (e.g., one disclosed in WO 2008/057637, in particular shown as SEQ ID NOs: 59 and 60).

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Thermoascus aurantiacus* AA9 (GH61A) polypeptide having cellulolytic enhancing activity (e.g., SEQ ID NO: 2 in WO 2005/074656), and *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499).

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Penicillium emersonii* AA9 (GH61A) polypeptide having cellulolytic enhancing activity, in particular the one disclosed in WO 2011/041397, and *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499).

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Penicillium emersonii* AA9 (GH61A) polypeptide having cellulolytic enhancing activity, in particular the one disclosed in WO 2011/041397, and *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499) or a variant disclosed in WO 2012/044915 (hereby incorporated by reference), in particular one comprising one or more such as all of the following substitutions: F100D, S283G, N456E, F512Y.

In an embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic composition, further comprising an AA9 (GH61A) polypeptide having cellulolytic enhancing activity, in particular the one derived from a strain of *Penicillium emersonii* (e.g., SEQ ID NO: 2 in WO 2011/041397), *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 in WO 2005/047499) variant with one or more, in particular all of the following substitutions: F100D, S283G, N456E, F512Y and disclosed in WO 2012/044915; *Aspergillus fumigatus* Cel7A CBH1, e.g., the one disclosed as SEQ ID NO: 6 in WO2011/057140 and *Aspergillus fumigatus* CBH II, e.g., the one disclosed as SEQ ID NO: 18 in WO 2011/057140.

In a preferred embodiment the cellulolytic enzyme composition is a *Trichoderma reesei,* cellulolytic enzyme composition, further comprising a hemicellulase or hemicellulolytic enzyme

CA 03158982 2022-5-19

composition, such as an *Aspergillus fumigatus* xylanase and *Aspergillus fumigatus* beta-xylosidase.

In an embodiment the cellulolytic enzyme composition also comprises a xylanase (e.g., derived from a strain of the genus *Aspergillus*, in particular *Aspergillus aculeatus* or *Aspergillus fumigatus;* or a strain of the genus *Talaromyces*, in particular *Talaromyces leycettanus*) and/or a beta-xylosidase (e.g., derived from *Aspergillus*, in particular *Aspergillus fumigatus*, or a strain of *Talaromyces*, in particular *Talaromyces emersonii*).

In an embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Thermoascus aurantiacus* AA9 (GH61A) polypeptide having cellulolytic enhancing activity (e.g., WO 2005/074656), *Aspergillus oryzae* beta-glucosidase fusion protein (e.g., one disclosed in WO 2008/057637, in particular as SEQ ID NOs: 59 and 60), and *Aspergillus aculeatus* xylanase (e.g., Xyl II in WO 94/21785).

In another embodiment the cellulolytic enzyme composition comprises a *Trichoderma reesei* cellulolytic preparation, further comprising *Thermoascus aurantiacus* GH61A polypeptide having cellulolytic enhancing activity (e.g., SEQ ID NO: 2 in WO 2005/074656), *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499) and *Aspergillus aculeatus* xylanase (Xyl II disclosed in WO 94/21785).

In another embodiment the cellulolytic enzyme composition comprises a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Thermoascus aurantiacus* AA9 (GH61A) polypeptide having cellulolytic enhancing activity (e.g., SEQ ID NO: 2 in WO 2005/074656), *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499) and *Aspergillus aculeatus* xylanase (e.g., Xyl II disclosed in WO 94/21785).

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Penicillium emersonii* AA9 (GH61A) polypeptide having cellulolytic enhancing activity, in particular the one disclosed in WO 2011/041397, *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499) and *Aspergillus fumigatus* xylanase (e.g., Xyl III in WO 2006/078256).

In another embodiment the cellulolytic enzyme composition comprises a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Penicillium emersonii* AA9 (GH61A) polypeptide having cellulolytic enhancing activity, in particular the one disclosed in WO 2011/041397, *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499), *Aspergillus fumigatus* xylanase (e.g., Xyl III in WO 2006/078256), and CBH I from *Aspergillus fumigatus*, in particular Cel7A CBH1 disclosed as SEQ ID NO: 2 in WO2011/057140.

CA 03158982 2022-5-19

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Penicillium emersonii* AA9 (GH61A) polypeptide having cellulolytic enhancing activity, in particular the one disclosed in WO 2011/041397, *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499), *Aspergillus fumigatus* xylanase (e.g., Xyl III in WO 2006/078256), CBH I from *Aspergillus fumigatus*, in particular Cel7A CBH1 disclosed as SEQ ID NO: 2 in WO 2011/057140, and CBH II derived from *Aspergillus fumigatus* in particular the one disclosed as SEQ ID NO: 4 in WO 2013/028928.

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Penicillium emersonii* AA9 (GH61A) polypeptide having cellulolytic enhancing activity, in particular the one disclosed in WO 2011/041397, *Aspergillus fumigatus* beta-glucosidase (e.g., SEQ ID NO: 2 of WO 2005/047499) or variant thereof with one or more, in particular all, of the following substitutions: F100D, S283G, N456E, F512Y; *Aspergillus fumigatus* xylanase (e.g., Xyl III in WO 2006/078256), CBH I from *Aspergillus fumigatus*, in particular Cel7A CBH I disclosed as SEQ ID NO: 2 in WO 2011/057140, and CBH II derived from *Aspergillus fumigatus*, in particular the one disclosed in WO 2013/028928.

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition comprising the CBH I (GENSEQP Accession No. AZY49536 (WO2012/103293); a CBH II (GENSEQP Accession No. AZY49446 (WO2012/103288); a beta-glucosidase variant (GENSEQP Accession No. AZU67153 (WO 2012/44915)), in particular with one or more, in particular all, of the following substitutions: F100D, S283G, N456E, F512Y; and AA9 (GH61 polypeptide) (GENSEQP Accession No. BAL61510 (WO 2013/028912)).

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition comprising a CBH I (GENSEQP Accession No. AZY49536 (WO2012/103293)); a CBH II (GENSEQP Accession No. AZY49446 (WO2012/103288); a GH10 xylanase (GENSEQP Accession No. BAK46118 (WO 2013/019827)); and a beta-xylosidase (GENSEQP Accession No. AZI04896 (WO 2011/057140)).

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition comprising a CBH I (GENSEQP Accession No. AZY49536 (WO2012/103293)); a CBH II (GENSEQP Accession No. AZY49446 (WO2012/103288)); and an AA9 (GH61 polypeptide; GENSEQP Accession No. BAL61510 (WO 2013/028912)).

In another embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition comprising a CBH I (GENSEQP Accession No. AZY49536

CA 03158982 2022-5-19

(WO2012/103293)); a CBH II (GENSEQP Accession No. AZY49446 (WO2012/103288)), an AA9 (GH61 polypeptide; GENSEQP Accession No. BAL61510 (WO 2013/028912)), and a catalase (GENSEQP Accession No. BAC11005 (WO 2012/130120)).

In an embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition comprising a CBH I (GENSEQP Accession No. AZY49446 (WO2012/103288); a CBH II (GENSEQP Accession No. AZY49446 (WO2012/103288)), a beta-glucosidase variant (GENSEQP Accession No. AZU67153 (WO 2012/44915)), with one or more, in particular all, of the following substitutions: F100D, S283G, N456E, F512Y; an AA9 (GH61 polypeptide; GENSEQP Accession No. BAL61510 (WO 2013/028912)), a GH10 xylanase (GENSEQP Accession No. BAK46118 (WO 2013/019827)), and a beta-xylosidase (GENSEQP Accession No. AZI04896 (WO 2011/057140)).

In an embodiment the cellulolytic composition is a *Trichoderma reesei* cellulolytic enzyme preparation comprising an EG I (Swissprot Accession No. P07981), EG II (EMBL Accession No. M19373), CBH I (*supra*); CBH II (*supra*); beta-glucosidase variant (*supra*) with the following substitutions: F100D, S283G, N456E, F512Y; an AA9 (GH61 polypeptide; *supra*), GH10 xylanase (*supra*); and beta-xylosidase (*supra*).

All cellulolytic enzyme compositions disclosed in WO 2013/028928 are also contemplated and hereby incorporated by reference.

The cellulolytic enzyme composition comprises or may further comprise one or more (several) proteins selected from the group consisting of a cellulase, a AA9 (i.e., GH61) polypeptide having cellulolytic enhancing activity, a hemicellulase, an expansin, an esterase, a laccase, a ligninolytic enzyme, a pectinase, a peroxidase, a protease, and a swollenin.

In one embodiment the cellulolytic enzyme composition is a commercial cellulolytic enzyme composition. Examples of commercial cellulolytic enzyme compositions suitable for use in a process of the invention include: CELLIC® CTec (Novozymes A/S), CELLIC® CTec2 (Novozymes A/S), CELLIC® CTec3 (Novozymes A/S), CELLUCLAST™ (Novozymes A/S), SPEZYME™ CP (Genencor Int.), ACCELLERASE™ 1000, ACCELLERASE 1500, ACCELLERASE™ TRIO (DuPont), FILTRASE® NL (DSM); METHAPLUS® S/L 100 (DSM), ROHAMENT™ 7069 W (Röhm GmbH), or ALTERNAFUEL® CMAX3™ (Dyadic International, Inc.). The cellulolytic enzyme composition may be added in an amount effective from about 0.001 to about 5.0 wt. % of solids, e.g., about 0.025 to about 4.0 wt. % of solids or about 0.005 to about 2.0 wt. % of solids.

Additional enzymes, and compositions thereof can be found in WO2011/153516 and WO2016/045569 (the contents of which are incorporated herein).

CA 03158982 2022-5-19

Additional polynucleotides encoding suitable cellulolytic enzymes may be obtained from microorganisms of any genus, including those readily available within the UniProtKB database (www.uniprot.org).

The cellulolytic enzyme coding sequences can also be used to design nucleic acid probes to identify and clone DNA encoding cellulolytic enzymes from strains of different genera or species, as described *supra*.

The polynucleotides encoding cellulolytic enzymes may also be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) as described *supra*.

Techniques used to isolate or clone polynucleotides encoding cellulolytic enzymes are described *supra*.

In one embodiment, the cellulolytic enzyme has a mature polypeptide sequence of at least 60%, e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any cellulolytic enzyme described or referenced herein (e.g., any endoglucanase, cellobiohydrolase, or beta-glucosidase). In one embodiment, the cellulolytic enzyme ha a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any cellulolytic enzyme described or referenced herein. In one embodiment, the cellulolytic enzyme has a mature polypeptide sequence that comprises or consists of the amino acid sequence of any cellulolytic enzyme described or referenced herein, allelic variant, or a fragment thereof having cellulolytic enzyme activity. In one embodiment, the cellulolytic enzyme has an amino acid substitution, deletion, and/or insertion of one or more (e.g., two, several) amino acids. In some embodiments, the total number of amino acid substitutions, deletions and/or insertions is not more than 10, e.g., not more than 9, 8, 7, 6, 5, 4, 3, 2, or 1.

In some embodiments, the cellulolytic enzyme has at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of the cellulolytic enzyme activity of any cellulolytic enzyme described or referenced herein (e.g., any endoglucanase, cellobiohydrolase, or beta-glucosidase) under the same conditions.

In one embodiment, the cellulolytic enzyme coding sequence hybridizes under at least low stringency conditions, e.g., medium stringency conditions, medium-high stringency conditions,

CA 03158982 2022-5-19

high stringency conditions, or very high stringency conditions with the full-length complementary strand of the coding sequence from any cellulolytic enzyme described or referenced herein (e.g., any endoglucanase, cellobiohydrolase, or beta-glucosidase). In one embodiment, the cellulolytic enzyme coding sequence has at least 65%, e.g., at least 70%, at least 75%, at least 80%, at least 85%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity with the coding sequence from any cellulolytic enzyme described or referenced herein.

In one embodiment, the polynucleotide encoding the cellulolytic enzyme comprises the coding sequence of any cellulolytic enzyme described or referenced herein (e.g., any endoglucanase, cellobiohydrolase, or beta-glucosidase). In one embodiment, the polynucleotide encoding the cellulolytic enzyme comprises a subsequence of the coding sequence from any cellulolytic enzyme described or referenced herein, wherein the subsequence encodes a polypeptide having cellulolytic enzyme activity. In one embodiment, the number of nucleotides residues in the subsequence is at least 75%, e.g., at least 80%, 85%, 90%, or 95% of the number of the referenced coding sequence.

The cellulolytic enzyme can also include fused polypeptides or cleavable fusion polypeptides, as described *supra*.

Fermentation products

A fermentation product can be any substance derived from the fermentation. The fermentation product can be, without limitation, an alcohol (e.g., arabinitol, n-butanol, isobutanol, ethanol, glycerol, methanol, ethylene glycol, 1,3-propanediol [propylene glycol], butanediol, glycerin, sorbitol, and xylitol); an alkane (e.g., pentane, hexane, heptane, octane, nonane, decane, undecane, and dodecane), a cycloalkane (e.g., cyclopentane, cyclohexane, cycloheptane, and cyclooctane), an alkene (e.g., pentene, hexene, heptene, and octene); an amino acid (e.g., aspartic acid, glutamic acid, glycine, lysine, serine, and threonine); a gas (e.g., methane, hydrogen (H_2), carbon dioxide (CO_2), and carbon monoxide (CO)); isoprene; a ketone (e.g., acetone); an organic acid (e.g., acetic acid, acetonic acid, adipic acid, ascorbic acid, citric acid, 2,5-diketo-D-gluconic acid, formic acid, fumaric acid, glucaric acid, gluconic acid, glucuronic acid, glutaric acid, 3-hydroxypropionic acid, itaconic acid, lactic acid, malic acid, malonic acid, oxalic acid, oxaloacetic acid, propionic acid, succinic acid, and xylonic acid); and polyketide.

In one embodiment, the fermentation product is an alcohol. The term "alcohol" encompasses a substance that contains one or more hydroxyl moieties. The alcohol can be, but is not limited to, n-butanol, isobutanol, ethanol, methanol, arabinitol, butanediol, ethylene glycol,

CA 03158982 2022-5-19

glycerin, glycerol, 1,3-propanediol, sorbitol, xylitol. See, for example, Gong *et al.*, 1999, Ethanol production from renewable resources, in *Advances in Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heidelberg, Germany, 65: 207-241; Silveira and Jonas, 2002, *Appl. Microbiol. Biotechnol.* 59: 400-408; Nigam and Singh, 1995, *Process Biochemistry* 30(2): 117-124; Ezeji *et al.*, 2003, *World Journal of Microbiology and Biotechnology* 19(6): 595-603. In one embodiment, the fermentation product is ethanol.

In another embodiment, the fermentation product is an alkane. The alkane may be an unbranched or a branched alkane. The alkane can be, but is not limited to, pentane, hexane, heptane, octane, nonane, decane, undecane, or dodecane.

In another embodiment, the fermentation product is a cycloalkane. The cycloalkane can be, but is not limited to, cyclopentane, cyclohexane, cycloheptane, or cyclooctane.

In another embodiment, the fermentation product is an alkene. The alkene may be an unbranched or a branched alkene. The alkene can be, but is not limited to, pentene, hexene, heptene, or octene.

In another embodiment, the fermentation product is an amino acid. The organic acid can be, but is not limited to, aspartic acid, glutamic acid, glycine, lysine, serine, or threonine. See, for example, Richard and Margaritis, 2004, *Biotechnology and Bioengineering* 87(4): 501-515.

In another embodiment, the fermentation product is a gas. The gas can be, but is not limited to, methane, H_2, CO_2, or CO. See, for example, Kataoka *et al.*, 1997, *Water Science and Technology* 36(6-7): 41-47; and Gunaseelan, 1997, *Biomass and Bioenergy* 13(1-2): 83-114.

In another embodiment, the fermentation product is isoprene.

In another embodiment, the fermentation product is a ketone. The term "ketone" encompasses a substance that contains one or more ketone moieties. The ketone can be, but is not limited to, acetone.

In another embodiment, the fermentation product is an organic acid. The organic acid can be, but is not limited to, acetic acid, acetonic acid, adipic acid, ascorbic acid, citric acid, 2,5-diketo-D-gluconic acid, formic acid, fumaric acid, glucaric acid, gluconic acid, glucuronic acid, glutaric acid, 3-hydroxypropionic acid, itaconic acid, lactic acid, malic acid, malonic acid, oxalic acid, propionic acid, succinic acid, or xylonic acid. See, for example, Chen and Lee, 1997, *Appl. Biochem. Biotechnol.* 63-65: 435-448.

In another embodiment, the fermentation product is polyketide.

In some embodiments, the host cell or fermenting organism (or processes thereof), provide higher yield of fermentation product (e.g., ethanol) when compared to the same cell without the heterologous polynucleotide encoding a sugar transporter described herein under the

CA 03158982 2022-5-19

same conditions (e.g., after 40 hours of fermentation). In some embodiments, the process results in at least 0.25%, such as 0.5%, 0.75%, 1.0%, 1.25%, 1.5%, 1.75%, 2%, 3% or 5% higher yield of the fermentation product (e.g., ethanol).

Recovery

The fermentation product, e.g., ethanol, can optionally be recovered from the fermentation medium using any method known in the art including, but not limited to, chromatography, electrophoretic procedures, differential solubility, distillation, or extraction. For example, alcohol is separated from the fermented cellulosic material and purified by conventional methods of distillation. Ethanol with a purity of up to about 96 vol. % can be obtained, which can be used as, for example, fuel ethanol, drinking ethanol, *i.e.*, potable neutral spirits, or industrial ethanol.

In some embodiments of the methods, the fermentation product after being recovered is substantially pure. With respect to the methods herein, "substantially pure" intends a recovered preparation that contains no more than 15% impurity, wherein impurity intends compounds other than the fermentation product (e.g., ethanol). In one variation, a substantially pure preparation is provided wherein the preparation contains no more than 25% impurity, or no more than 20% impurity, or no more than 10% impurity, or no more than 5% impurity, or no more than 3% impurity, or no more than 1% impurity, or no more than 0.5% impurity.

Suitable assays to test for the production of ethanol and contaminants, and sugar consumption can be performed using methods known in the art. For example, ethanol product, as well as other organic compounds, can be analyzed by methods such as HPLC (High Performance Liquid Chromatography), GC-MS (Gas Chromatography Mass Spectroscopy) and LC-MS (Liquid Chromatography-Mass Spectroscopy) or other suitable analytical methods using routine procedures well known in the art. The release of ethanol in the fermentation broth can also be tested with the culture supernatant. Byproducts and residual sugar in the fermentation medium (e.g., glucose or xylose) can be quantified by HPLC using, for example, a refractive index detector for glucose and alcohols, and a UV detector for organic acids (Lin et al., *Biotechnol. Bioeng.* 90:775 -779 (2005)), or using other suitable assay and detection methods well known in the art.

The invention described and claimed herein is not to be limited in scope by the specific aspects or embodiments herein disclosed, since these aspects/embodiments are intended as illustrations of several aspects of the invention. Any equivalent aspects are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those

CA 03158982 2022-5-19

shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. In the case of conflict, the present disclosure including definitions will control. All references are specifically incorporated by reference for that which is described.

The following examples are offered to illustrate certain aspects/embodiments of the present invention, but not in any way intended to limit the scope of the invention as claimed.

The invention may further be described in the following numbered paragraphs:

Paragraph [1]. A recombinant host cell comprising a heterologous polynucleotide encoding a sugar transporter, wherein the transporter has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 257-397; and wherein the cell comprises an active pentose fermentation pathway.

Paragraph [2]. The recombinant host cell of paragraph [1], wherein the heterologous polynucleotide encoding a sugar transporter is operably linked to a promoter that is foreign to the polynucleotide.

Paragraph [3]. The recombinant host cell of paragraph [1] or [2], wherein the heterologous polynucleotide encodes a sugar transporter having a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any one of SEQ ID NOs: 257-397.

Paragraph [4]. The recombinant host cell of any one of paragraphs [1]-[3], wherein the heterologous polynucleotide encodes a sugar transporter has a mature polypeptide sequence comprising or consisting of the amino acid sequence of any one of SEQ ID NOs: 257-397.

Paragraph [5]. The recombinant host cell of paragraphs [1]-[4], wherein the cell comprises an active xylose fermentation pathway.

Paragraph [6]. The recombinant host cell of paragraph [5], wherein the cell comprises one or more active xylose fermentation pathway genes selected from:
a heterologous polynucleotide encoding a xylose isomerase (XI), and

a heterologous polynucleotide encoding a xylulokinase (XK).

Paragraph [7]. The recombinant host cell of paragraph [5] or [6], wherein the cell comprises one or more active xylose fermentation pathway genes selected from:
a heterologous polynucleotide encoding a xylose reductase (XR),
a heterologous polynucleotide encoding a xylitol dehydrogenase (XDH), and
a heterologous polynucleotide encoding a xylulokinase (XK).

Paragraph [8]. The recombinant host cell of any one of paragraphs [1]-[7], wherein the cell comprises an active arabinose fermentation pathway.

Paragraph [9]. The recombinant host cell of paragraph [8], wherein the cell comprises one or more active arabinose fermentation pathway genes selected from:
a heterologous polynucleotide encoding a L-arabinose isomerase (AI),
a heterologous polynucleotide encoding a L-ribulokinase (RK), and
a heterologous polynucleotide encoding a L-ribulose-5-P4-epimerase (R5PE).

Paragraph [10]. The recombinant host cell of paragraph [8] or [9], wherein the cell comprises one or more active arabinose fermentation pathway genes selected from:
a heterologous polynucleotide encoding an aldose reductase (AR),
a heterologous polynucleotide encoding a L-arabinitol 4-dehydrogenase (LAD),
a heterologous polynucleotide encoding a L-xylulose reductase (LXR),
a heterologous polynucleotide encoding a xylitol dehydrogenase (XDH) and
a heterologous polynucleotide encoding a xylulokinase (XK).

Paragraph [11]. The recombinant host cell of any one of paragraphs [1]-[10], the cell comprises an active xylose fermentation pathway and an active arabinose fermentation pathway.

Paragraph [12]. The recombinant host cell of any one of paragraphs [1]-[11], with the proviso that the sugar transporter is not the transporter having a mature polypeptide sequence of SEQ ID NO: 390 (or a transporter having a mature polypeptide sequence with at least 80%, e.g., at least 85%, 90%, 95%, 97%, 98%, or 99% sequence identity to the transporter of SEQ ID NO: 390).

Paragraph [13]. A recombinant host cell comprising a heterologous polynucleotide encoding a sugar transporter, wherein the transporter has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity

to any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131 and wherein the cell comprises an active arabinose fermentation pathway.

Paragraph [14]. The recombinant host cell of paragraph [13], wherein the heterologous polynucleotide encodes a sugar transporter having a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131.

Paragraph [15]. The recombinant host cell of paragraph [13] or [14], wherein the heterologous polynucleotide encodes a sugar transporter having a mature polypeptide sequence comprising or consisting of the amino acid sequence of any one of SEQ ID NOs: 40, 53, 63, 72, 99, 108, 111, 123, 124 and 131.

Paragraph [16]. A recombinant host cell comprising a heterologous polynucleotide encoding a sugar transporter, wherein the transporter has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 97, 116 and 138 and wherein the cell comprises an active xylose fermentation pathway.

Paragraph [17]. The recombinant host cell of paragraph [16], wherein the heterologous polynucleotide encodes a sugar transporter with a mature polypeptide sequence that differs by no more than ten amino acids, e.g., by no more than five amino acids, by no more than four amino acids, by no more than three amino acids, by no more than two amino acids, or by one amino acid from any one of SEQ ID NOs: 97, 116 and 138.

Paragraph [18]. The recombinant host cell of paragraph [16] or [17], wherein the heterologous polynucleotide encodes a sugar transporter having a mature polypeptide sequence comprising or consisting of the amino acid sequence of any one of SEQ ID NOs: 97, 116 and 138.

Paragraph [19]. The recombinant host cell of any one of paragraphs [1]-[18], wherein the cell further comprises a heterologous polynucleotide encoding a glucoamylase.

Paragraph [20]. The recombinant host cell of paragraph [19], wherein the glucoamylase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 8, 102-113, 229, 230 and 244-250.

CA 03158982 2022-5-19

Paragraph [21]. The recombinant host cell of paragraph [19] or [20], wherein the heterologous polynucleotide encoding the glucoamylase is operably linked to a promoter that is foreign to the polynucleotide.

Paragraph [22]. The recombinant host cell of any one of paragraphs [1]-[21], wherein the cell further comprises a heterologous polynucleotide encoding an alpha-amylase.

Paragraph [23]. The recombinant host cell of paragraph [22], wherein the alpha-amylase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 76-101, 121-174, 231 and 251-256.

Paragraph [24]. The recombinant host cell of paragraph [22] or [23], wherein the heterologous polynucleotide encoding the alpha-amylase is operably linked to a promoter that is foreign to the polynucleotide.

Paragraph [25]. The recombinant host cell of any one of paragraphs [1]-[24], wherein the cell further comprises a heterologous polynucleotide encoding a phospholipase.

Paragraph [26]. The recombinant host cell of paragraph [25], wherein the phospholipase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 235, 236, 237, 238, 239, 240, 241 and 242.

Paragraph [27]. The recombinant host cell of paragraph [25] or [26], wherein the heterologous polynucleotide encoding phospholipase is operably linked to a promoter that is foreign to the polynucleotide.

Paragraph [28]. The recombinant host cell of any one of paragraphs [1]-[27], wherein the cell further comprises a heterologous polynucleotide encoding a trehalase.

Paragraph [29]. The recombinant host cell of paragraph [28], wherein the trehalase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 175-226.

CA 03158982 2022-5-19

Paragraph [30]. The recombinant host cell of paragraph [27] or [28], wherein the heterologous polynucleotide encoding the trehalase is operably linked to a promoter that is foreign to the polynucleotide.

Paragraph [31]. The recombinant host cell of any one of paragraphs [1]-[30], wherein the cell further comprises a heterologous polynucleotide encoding a protease.

Paragraph [32]. The recombinant host cell of paragraph [31], wherein the protease has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 9-73.

Paragraph [33]. The recombinant host cell of paragraph [31] or [32], wherein the heterologous polynucleotide encoding the protease is operably linked to a promoter that is foreign to the polynucleotide.

Paragraph [34]. The recombinant host cell of any one of paragraphs [1]-[33], wherein the cell further comprises a heterologous polynucleotide encoding a pullulanase.

Paragraph [35]. The recombinant host cell of paragraph [34], wherein the pullulanase has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity the amino acid sequence of any one of SEQ ID NOs: 114-120.

Paragraph [36]. The recombinant host cell of paragraph [34] or [35], wherein the heterologous polynucleotide encoding the pullulanase is operably linked to a promoter that is foreign to the polynucleotide.

Paragraph [37]. The recombinant host cell of any one of paragraphs [1]-[36], wherein the cell is capable of higher anaerobic growth rate on pentose (e.g., xylose and/or arabinose) compared to the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2).

Paragraph [38]. The recombinant host cell of any one of paragraphs [1]-[38], wherein the cell is capable of a higher rate of pentose consumption (e.g., at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 75% or 90% higher xylose and/or arabinose consumption) compared

to the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2).

Paragraph [39]. The recombinant host cell of any one of paragraphs [1]-[38], wherein the cell is capable of higher pentose (e.g., xylose and/or arabinose) consumption compared to the same cell without the heterologous polynucleotide encoding a sugar transporter at about or after 120 hours fermentation (e.g., under conditions described in Example 2).

Paragraph [40]. The recombinant host cell of paragraph [39], wherein the cell is capable of consuming more than 65%, e.g., at least 70%, 75%, 80%, 85%, 90%, 95% of pentose (e.g., xylose and/or arabinose) in the medium at about or after 120 hours fermentation (e.g., under conditions described in Example 2).

Paragraph [41]. The recombinant host cell of any one of paragraphs [1]-[40], wherein the cell is capable of higher ethanol production compared to the same cell without the heterologous polynucleotide encoding a sugar transporter under the same conditions (e.g., after 40 hours of fermentation).

Paragraph [42]. The recombinant host cell of any one of paragraphs [1]-[41], wherein the cell further comprises a heterologous polynucleotide encoding a transketolase (TKL1).

Paragraph [43]. The recombinant host cell of any one of paragraphs [1]-[42], wherein the cell further comprises a heterologous polynucleotide encoding a transaldolase (TAL1).

Paragraph [44]. The recombinant host cell of any one of paragraphs [1]-[43], wherein the cell further comprises a disruption to an endogenous gene encoding a glycerol 3-phosphate dehydrogenase (GPD).

Paragraph [45]. The recombinant host cell of any one of paragraphs [1]-[44], wherein the cell further comprises a disruption to an endogenous gene encoding a glycerol 3-phosphatase (GPP).

Paragraph [46]. The recombinant host cell of any one of paragraphs [1]-[45], wherein the cell is a yeast cell.

Paragraph [47]. The recombinant host cell of any one of paragraphs [1]-[46], wherein the cell is a *Saccharomyces, Rhodotorula, Schizosaccharomyces, Kluyveromyces, Pichia, Hansenula, Rhodosporidium, Candida, Yarrowia, Lipomyces, Cryptococcus,* or *Dekkera sp.* cell.

CA 03158982 2022-5-19

Paragraph [48]. The recombinant host cell of any one of paragraphs [1]-[47], wherein the cell is a *Saccharomyces cerevisiae* cell.

Paragraph [49]. A composition comprising the recombinant host cell of any one of paragraphs [1]-[48] and one or more naturally occurring and/or non-naturally occurring components, such as components are selected from the group consisting of: surfactants, emulsifiers, gums, swelling agents, and antioxidants.

Paragraph [50]. A method of producing a derivative of a recombinant host cell of any one of paragraphs [1]-[49], the method comprising:

(a) providing:

(i) a first host cell; and

(ii) a second host cell, wherein the second host cell is a recombinant host cell of any one of paragraphs [1]-[49];

(b) culturing the first host cell and the second host cell under conditions which permit combining of DNA between the first and second host cells;

(c) screening or selecting for a derive host cell.

Paragraph [51]. A method of producing a fermentation product from a starch-containing or cellulosic-containing material, the method comprising:

(a) saccharifying the starch-containing or cellulosic-containing material; and

(b) fermenting the saccharified material of step (a) with the recombinant host cell of any one of paragraphs [1]-[50] under suitable conditions to produce the fermentation product.

Paragraph [52]. The method of paragraph [51], wherein saccharification of step (a) occurs on a starch-containing material, and wherein the starch-containing material is either gelatinized or ungelatinized starch.

Paragraph [53]. The method of paragraph [52], comprising liquefying the starch-containing material by contacting the material with an alpha-amylase prior to saccharification.

Paragraph [54]. The method of paragraph [52] or [53], wherein liquefying the starch-containing material and/or saccharifying the starch-containing material is conducted in presence of exogenously added protease.

Paragraph [55]. The method of any one of paragraphs [51]-[54], wherein fermentation is performed under reduced nitrogen conditions (e.g., less than 1000 ppm urea or ammonium

CA 03158982 2022-5-19

hydroxide, such as less than 750 ppm, less than 500 ppm, less than 400 ppm, less than 300 ppm, less than 250 ppm, less than 200 ppm, less than 150 ppm, less than 100 ppm, less than 75 ppm, less than 50 ppm, less than 25 ppm, or less than 10 ppm).

Paragraph [56]. The method of any one of paragraphs [51]-[55], wherein fermentation and saccharification are performed simultaneously in a simultaneous saccharification and fermentation (SSF).

Paragraph [57]. The method of any one of paragraphs [51]-[55], wherein fermentation and saccharification are performed sequentially (SHF).

Paragraph [58]. The method of any one of paragraphs paragraph [51]-[57], comprising recovering the fermentation product from the fermentation.

Paragraph [59]. The method of paragraph [58], wherein recovering the fermentation product from the fermentation comprises distillation.

Paragraph [60]. The method of any one of paragraphs [51]-[59], wherein the fermentation product is ethanol.

Paragraph [61]. The method of any one of paragraphs [51]-[60], wherein step (a) comprises contacting the cellulosic and/or starch-containing with an enzyme composition.

Paragraph [62]. The method of any one of paragraphs [51]-[61], wherein saccharification occurs on a cellulosic material, and wherein the cellulosic material is pretreated.

Paragraph [63]. The method of paragraph [62], wherein the pretreatment is a dilute acid pretreatment.

Paragraph [64]. The method of paragraph [62] or [63], wherein saccharification occurs on a cellulosic material, and wherein step (a) comprises contacting the cellulosic enzyme composition, and wherein the enzyme composition comprises one or more enzymes selected from a cellulase, an AA9 polypeptide, a hemicellulase, a CIP, an esterase, an expansin, a ligninolytic enzyme, an oxidoreductase, a pectinase, a protease, and a swollenin.

Paragraph [65]. The method of paragraph [64], wherein the cellulase is one or more enzymes selected from an endoglucanase, a cellobiohydrolase, and a beta-glucosidase.

CA 03158982 2022-5-19

Paragraph [66]. The method of paragraph [64] or [65], wherein the hemicellulase is one or more enzymes selected a xylanase, an acetylxylan esterase, a feruloyl esterase, an arabinofuranosidase, a xylosidase, and a glucuronidase.

Paragraph [67]. The method of any one of paragraphs [51]-[66], wherein the method results in higher yield of fermentation product when compared to the method using the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2).

Paragraph [68]. The method of paragraph [67], wherein the method results in at least 0.25% (e.g., 0.5%, 0.75%, 1.0%, 1.25%, 1.5%, 1.75%, 2%, 3% or 5%) higher yield of fermentation product.

Paragraph [69]. The method of any one of paragraphs [51]-[68], wherein fermentation is conducted under low oxygen (e.g., anaerobic) conditions.

Paragraph [70]. The method of any one of paragraphs [51]-[69] wherein a greater amount of pentose (e.g., xylose and/or arabinose) is consumed (e.g., at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 60%, 75% or 90% more) when compared to the method using the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2).

Paragraph [71]. The method of any one of any one of paragraphs [51]-[70], wherein more than 65%, e.g., at least 70%, 75%, 80%, 85%, 90%, 95% of pentose (e.g., xylose and/or arabinose) in the medium is consumed (e.g., under conditions described in Example 2).

Paragraph [72]. Use of a recombinant host cell of any one of paragraphs [1]-[49] in the production of ethanol.

Examples

Materials and Methods

Chemicals used as buffers and substrates were commercial products of at least reagent grade.

Yeast strain S509-C04 is was prepared according the breeding procedures described in US Patent No. 8,257,959 and further comprises an active arabinose and xylose fermentation pathways with heterologous genes expressing L-arabinitol 4-dehydrogenase (LAD), L-xylulose reductase (LXR), D-xylulose reductase xylitol dehydrogenase (XDH) and xylulokinase (XK). *See*,

CA 03158982 2022-5-19

U.S. Provisional Application Number 63/024,010 entitled "Improved Microorganisms for Arabinose Fermentation" filed May 13, 2020.

Example 1: Construction of yeast strains expressing a heterologous sugar transporter

This example describes the construction of yeast cells containing a heterologous sugar transporter under the control of an *S. cerevisiae* TEF2 promoter (SEQ ID NO: 2). Three or four pieces of DNA containing the promoter, gene (or gene in two pieces) and terminator were designed to allow for homologous recombination between the 3 or 4 DNA fragments and into the Int1 locus of the yeast strain S509-C04. The resulting strains have one fragment (left fragment) containing the TEF2 promoter, one fragment (middle fragment) containing the sugar transporter coding sequence and one fragment (right fragment) containing the PRM9 terminator (SEQ ID NO: 243) integrated into the *S. cerevisiae* genome at the XII-2 locus.

Construction of the promoter-containing fragments (left fragments)

Synthetic linear uncloned DNA containing 300 bp homology to the Int1 site and the *S. cerevisiae* TEF2 promoter was synthesized by Thermo Fischer Scientific (Waltham, MA) and PCR-amplified with primers 1229212 (5'-TTCTT ACCAA TCCTT TCATA-3'; SEQ ID NO: 398) and 1224106 (5'-TTTGT TCTAG CTTAA TTATA GTTCG TTG-3'; SEQ ID NO: 399). Fifty pmoles each of forward and reverse primer was used in 24 PCR reactions containing 100 ng of plasmid DNA as template and 5X Platinum SuperFi PCR Master Mix (Thermo Fisher Scientific) in a final volume of 50 µL. The PCR was performed in a T100™ Thermal Cycler (Bio-Rad Laboratories, Inc.; Hercules, CA) programmed for one cycle at 98°C for 30 seconds followed by 30 cycles each at 98°C for 10 seconds, 54°C for 10 seconds, and 72°C for 30 seconds with a final extension at 72°C for 5 minutes. Following thermocycling, the PCR reaction products gel isolated and cleaned up using the NucleoSpin Gel and PCR clean-up kit (Machery-Nagel; Düren, Germany).

Construction of the transporter-containing fragments (middle fragments)

Synthetic linear uncloned DNA containing 50 bp of the TEF2 promoter, the sugar transporter gene and 50bp of the PRM9 terminator were synthesized by Twist Bioscience (San Francisco, CA) or Thermo Fisher Scientific.

Construction of the terminator-contain fragment (right fragment)

Synthetic linear uncloned DNA 18AFCGPC containing the *S. cerevisiae* PRM9 terminator and 300 bp homology to the Int1 site was synthesized by Thermo Fisher Scientific and PCR-amplified with primers 1229631 (5'-TAAAC AGAAG ACGGG AGACA CTAG-3'; SEQ ID NO: 400) and 1229213 (5'-AGGGC TAAAG TCTCA TGAAA-3'; SEQ ID NO: 401). Fifty pmoles each of

CA 03158982 2022-5-19

forward and reverse primer was used in 24 PCR reactions containing 100 ng of plasmid DNA as template and 5X Platinum SuperFi PCR Master Mix (Thermo Fisher Scientific) in a final volume of 50 µL. The PCR was performed in a T100™ Thermal Cycler (Bio-Rad Laboratories, Inc.) programmed for one cycle at 98°C for 30 seconds followed by 30 cycles each at 98°C for 10 seconds, 59°C for 10 seconds, and 72°C for 30 seconds with a final extension at 72°C for 5 minutes. Following thermocycling, the PCR reaction products gel isolated and cleaned up using the NucleoSpin Gel and PCR clean-up kit (Machery-Nagel).

Integration of the left, middle and right-hand fragments

The yeast strain S509-C04 was transformed with the left, middle and right integration fragments described above. In each transformation pool a fixed left fragment and right fragment with 200ng of each fragment was used. The middle fragment(s) consisted of the sugar transporter gene with 200-400ng of each fragment. To aid homologous recombination of the left, middle and right fragments at the genomic Int1 sites a plasmid containing cas9 and guide RNA specific to XII-2 (pMIBa457; Figure 3) was also used in the transformation. These 4-5 components were transformed into the into *S. cerevisiae* strain S509-C04 following a yeast electroporation protocol. Transformants were selected on YPD+cloNAT to select for transformants that contain the cas9 plasmid pMIBa457. Transformants were picked using a Q-pix Colony Picking System (Molecular Devices; San Jose, CA) to inoculate 1 well of 96-well plate containing YPD+cloNAT media. The plates were grown for 2 days then glycerol was added to 20% final concentration and the plates were stored at -80°C until needed. Integration of specific sugar transporter construct was verified by PCR with locus specific primers and subsequent sequencing. The resulting strains were used in the following examples as described below.

Example 2: Evaluation of yeast strains expressing a heterologous sugar transporter

Yeast strains from Example 1 expressing a heterologous sugar transporter were evaluated for growth in media where xylose or arabinose were the sole carbon source. The Growth Profiler (Enzyscreen; Heemstede, Netherlands) was used to evaluate strain growth. The Growth Profiler is an incubator that can simultaneously control growth conditions, take images of clear-bottom multi-titer growth plates, and measure cell density over time. The software GP Viewer converts pixels of defined regions per well of each image to RBG (red, blue, green) values; green values are translated to identify growth rates for analysis.

To prepare the strains for evaluation of growth in YNB+0.5% arabinose or 0.5% xylose media, yeast strains were grown for 24 hours in YPD medium with 2% glucose, at 30°C and 300

RPM. An inoculum of yeast was added to Growth Profiler plates containing 250uL of medium (YNB with 0.5% arabinose or 0.5% xylose). Plates are secured in the Growth Profiler and grown at 250 RPM, 30°C for 120 hours. Time intervals between each photo was 10 minutes. Growth evaluation was quenched by adding and mixing 50uL of 8% H_2SO_4. Samples were centrifuged at 3000 RPM for 10 minutes and the supernatant was collected for HPLC analysis for remaining arabinose and xylose concentrations. Slope of each strain was calculated by taking the ratio of rise (green value) over run (time (hours)) during exponential phase. Strains with the highest slopes were able to grow best in the media and those with the least amount of remaining arabinose or xylose consumed the most C5 sugar. Results are shown in Table 7.

Table 7.

Strain ID	Gene ID	Gene DB	Donor Organism	Arabinose Slope (g-value/h)	Arabinose consumed (g/L)	Xylose Slope (g-value/h)	Xylose consumed (g/L)
S509-C04	-----	-----	-----	0.2758737	3.08525	0.522	4.71275
S622-A05	A1C8W7	SWISSPROT	Aspergillus clavatus	0.0002404	0.596	0	0.05
S622-A10	A0A0R1SSI1	SWISSPROT	Lactobacillus versmoldensis	0.2765996	2.984	0.516	4.694
S622-B10	A0A0R1SSI1	SWISSPROT	Lactobacillus versmoldensis	0.3355385	3.272	0.517	4.664
S622-C01	A0A078DBU3	SWISSPROT	Brassica napus	0.396221	4.916	0.477	4.712
S622-C02	A0A087HLR1	SWISSPROT	Arabis alpina	0.3901352	3.548	0.523	4.682
S622-C08	A1C8W7	SWISSPROT	Aspergillus clavatus	-0.004996	0.62	0	0
S622-C10	A0A0R1SSI1	SWISSPROT	Lactobacillus versmoldensis	0.3364993	3.176	0.436	4.406
S622-C12	A0A0W0DYZ4	TREMBL	Candida glabrata (Torulopsis glabrata)	0.2900236	2.96	0.513	4.682
S622-D01	A0A078DBU3	SWISSPROT	Brassica napus	0.4051435	4.94	0.398	4.37
S622-D02	A0A087HLR1	SWISSPROT	Arabis alpina	0.3841702	3.404	0.487	4.652
S622-D07	A1C8W7	SWISSPROT	Aspergillus clavatus	0.6346935	5	0.543	5
S622-D08	A1DAC2	SWISSPROT	Neosartorya fischeri (Aspergillus fischerianus)	0.4000574	3.236	0.531	4.712
S622-D10	A0A0R1SSI1	SWISSPROT	Lactobacillus versmoldensis	0.2935952	2.972	0.459	4.334
S622-D12	A0A0W0DYZ4	TREMBL	Candida glabrata (Torulopsis glabrata)	0.260369	2.864	0.524	4.676
S622-E01	A0A078DBU3	SWISSPROT	Brassica napus	0.3742188	4.508	0.418	4.322

CA 03158982 2022-5-19

S622-E03	A0A090BHJ4	SWISSPROT	Kluyveromyces marxianus (Candida kefyr)	0.4265329	3.908	0.503	4.736
S622-E04	P43581	SWISSPROT	Saccharomyces cerevisiae	0.3291398	3.128	0.53	4.718
S622-E08	A1DAC2	SWISSPROT	Neosartorya fischeri (Aspergillus fischerianus)	0.3303353	3.284	0.569	4.778
S622-F03	A0A090BHJ4	SWISSPROT	Kluyveromyces marxianus (Candida kefyr)	0.4234165	3.836	0.54	4.754
S622-F04	P43581	SWISSPROT	Saccharomyces cerevisiae	0.3804804	3.116	0.489	4.67
S622-F07	A1C8W7	SWISSPROT	Aspergillus clavatus	0.519784	5	0.526	4.802
S622-F10	A0A0R1SSI1	SWISSPROT	Lactobacillus versmoldensis	0.3267609	3.212	0.557	4.67
S622-G02	A0A087HLR1	SWISSPROT	Arabis alpina	0.3307279	3.308	0.433	4.388
S622-G03	A0A090BHJ4	SWISSPROT	Kluyveromyces marxianus (Candida kefyr)	0.4192362	3.86	0.537	4.64
S622-G04	P43581	SWISSPROT	Saccharomyces cerevisiae	0.3282136	2.996	0.563	4.7
S622-G10	A0A0R1SSI1	SWISSPROT	Lactobacillus versmoldensis	0.3147277	3.056	0.551	4.61
S622-H03	A0A090BHJ4	SWISSPROT	Kluyveromyces marxianus (Candida kefyr)	0.1847741	2.864	0.539	4.658
S622-H04	P43581	SWISSPROT	Saccharomyces cerevisiae	0.3771452	3.392	0.535	4.664
S622-H07	A1C8W7	SWISSPROT	Aspergillus clavatus	0.2906293	4.412	0.515	4.784
S622-H08	A1DAC2	SWISSPROT	Neosartorya fischeri (Aspergillus fischerianus)	0.312139	3.248	0.563	4.742
S622-H12	A0A0R1SSI1	SWISSPROT	Lactobacillus versmoldensis	0.3497422	3.224	0.585	4.664
S623-A04	AOD63543	GENESEQP	Pichia guilliermondii	0.3838818	3.716	0.55	4.778
S623-A05	A0A1A0HK54	SWISSPROT	Metschnikowia bicuspidata var. bicuspidata	0.2624329	2.84	0.532	4.712
S623-A06	A0A1E4T6F0	TREMBL	Candida arabinofermentans	0.3354888	3.164	0.514	4.76
S623-A11	B6HE12	SWISSPROT	Penicillium rubens (Penicillium chrysogenum)	0.4258612	4.844	0.489	4.778
S623-A12	BBZ79998	GENESEQP	Candida intermedia	0.3688479	3.38	0	4.646

S623-B02	A5DWD7	SWISSPROT	Lodderomyces elongisporus (Saccharomyces elongisporus)	0.3527794	3.836	0.551	5
S623-B05	A0A1A0HK54	SWISSPROT	Metschnikowia bicuspidata var. bicuspidata	0.3198046	3.164	0.516	4.736
S623-B08	A0A1G4MFR0	SWISSPROT	Lachancea fermentati (Zygosaccharomyces fermentati)	0.4174188	3.956	0.488	4.796
S623-B09	B1H0U6	SWISSPROT	Ambrosiozyma monospora	0.3155448	2.984	0.507	4.658
S623-C02	A5DWD7	SWISSPROT	Lodderomyces elongisporus (Saccharomyces elongisporus)	0.4013034	3.968	0.517	4.79
S623-C03	ABN64726	GENPEPT	Pichia stipitis	0.290072	3.512	0.523	4.676
S623-C04	AOD63543	GENESEQP	Pichia guilliermondii	0.3696066	3.812	0.532	4.664
S623-C05	A0A1A0HK54	SWISSPROT	Metschnikowia bicuspidata var. bicuspidata	0.3195019	3.152	0.53	4.634
S623-C10	B1H0U7	SWISSPROT	Ambrosiozyma monospora	0.3464645	2.756	0.537	4.436
S623-D03	ABN64726	GENPEPT	Pichia stipitis	0.3427166	3.644	0.545	4.754
S623-D07	A0A1G4JE77	TREMBL	Lachancea meyersii	0.2694658	3.02	0.507	4.58
S623-D08	A0A1G4MFR0	SWISSPROT	Lachancea fermentati (Zygosaccharomyces fermentati)	0.456073	3.908	0.523	4.694
S623-D09	B1H0U6	SWISSPROT	Ambrosiozyma monospora	0.4450422	5	0.433	4.58
S623-D10	B1H0U7	SWISSPROT	Ambrosiozyma monospora	0.3692674	3.032	0.539	4.364
S623-D12	BBZ79998	GENESEQP	Candida intermedia	0.3983482	3.704	0.566	4.298
S623-E02	A5DWD7	SWISSPROT	Lodderomyces elongisporus (Saccharomyces elongisporus)	0.393257	3.884	0.495	4.778
S623-E07	A0A1G4JE77	TREMBL	Lachancea meyersii	0.2756948	3.08	0.478	4.784
S623-E08	A0A1G4MFR0	SWISSPROT	Lachancea fermentati (Zygosaccharomyces fermentati)	0.3490631	4.136	0.487	4.796
S623-E10	B1H0U7	SWISSPROT	Ambrosiozyma monospora	0.4128228	3.476	0.479	4.724
S623-F01	A5DPY9	SWISSPROT	Meyerozyma guilliermondii	0.3230665	3.176	0.563	4.778

CA 03158982 2022-5-19

			(Candida guilliermondii)				
S623-F02	A5DWD7	SWISSPROT	Lodderomyces elongisporus (Saccharomyces elongisporus)	0.4153273	3.788	0.451	4.394
S623-F03	ABN64726	GENPEPT	Pichia stipitis	0.3152494	3.56	0.511	4.7
S623-F04	AOD63543	GENESEQP	Pichia guilliermondii	0.2754059	3.068	0.571	4.688
S623-F07	A0A1G4JE77	TREMBL	Lachancea meyersii	0.3071527	2.9	0.477	4.502
S623-G05	A0A1A0HK54	SWISSPROT	Metschnikowia bicuspidata var. bicuspidata	0.3063656	3.152	0.513	4.688
S623-G06	A0A1E4T6F0	TREMBL	Candida arabinofermentans	0.3439764	3.128	0.541	4.646
S623-G09	B1H0U6	SWISSPROT	Ambrosiozyma monospora	0.2763948	3.968	0.503	4.742
S623-H02	A5DWD7	SWISSPROT	Lodderomyces elongisporus (Saccharomyces elongisporus)	0.3363999	3.776	0.553	4.826
S623-H03	ABN64726	GENPEPT	Pichia stipitis	0.4032777	3.584	0.507	4.646
S623-H04	AOD63543	GENESEQP	Pichia guilliermondii	0.3763735	4.28	0.566	4.784
S623-H05	A0A1A0HK54	SWISSPROT	Metschnikowia bicuspidata var. bicuspidata	0.3924578	3.368	0.538	4.58
S623-H07	A0A1G4JE77	TREMBL	Lachancea meyersii	0.3109583	3.068	0.418	4.766
S623-H10	B1H0U7	SWISSPROT	Ambrosiozyma monospora	0.3533139	2.708	0.53	4.472
S623-H12	BBZ79998	GENESEQP	Candida intermedia	0.4142323	3.704	0.559	4.688
S624-A01	A0A1K2I4N4	SWISSPROT	Lactobacillus rennini	0.3198507	4.856	0.508	4.55
S624-A02	A0A1L0BAU2	SWISSPROT	Candida intermedia	0.2888697	2.876	0.543	4.364
S624-A06	BFJ94472	GENESEQP	Metschnikowia sp	0.3889313	3.2	0.531	4.724
S624-A10	A0A1N6MBZ0	SWISSPROT	Candida galli	0.381647	3.344	0.571	4.31
S624-B02	A0A1L0BAU2	SWISSPROT	Candida intermedia	0.2858605	2.888	0.519	4.718
S624-B05	BFJ89633	GENESEQP	Saccharomyces cerevisiae	0.3280016	3.044	0.557	4.454
S624-B06	BFJ94472	GENESEQP	Metschnikowia sp	0.3540393	3.164	0.513	4.724
S624-C01	A0A1K2I4N4	SWISSPROT	Lactobacillus rennini	0.311077	2.828	0.526	4.676
S624-C03	A0A1L0BZU1	SWISSPROT	Candida intermedia	0.3274216	2.624	0.531	3.944

S624-C05	BFJ89633	GENESEQP	Saccharomyces cerevisiae	0.3350908	3.044	0.551	4.604
S624-C06	BFJ94472	GENESEQP	Metschnikowia sp	0.3454849	2.924	0.524	4.766
S624-C07	BFJ94474	GENESEQP	Metschnikowia sp	0.2834116	2.576	0.462	4.724
S624-C10	A0A1N6MBZ0	SWISSPROT	Candida galli	0.3861381	3.296	0.566	4.502
S624-C12	A0A1Y6JY60	SWISSPROT	Lactobacillus zymae	0.3042494	2.852	0.514	4.508
S624-D08	C4B4V9	SWISSPROT	Corynebacterium glutamicum (Brevibacterium saccharolyticum)	0.3458725	2.876	0.502	4.364
S624-D12	A0A1Y6JY60	SWISSPROT	Lactobacillus zymae	0.3330445	2.972	0.437	4.136
S624-E02	A0A1L0BAU2	SWISSPROT	Candida intermedia	0.3179369	3.008	0.515	4.778
S624-E04	A0A1L9U9S9	SWISSPROT	Aspergillus brasiliensis	0.376655	4.916	0.436	4.616
S624-E07	BFJ94474	GENESEQP	Metschnikowia sp	0.1852056	2.444	0.58	4.664
S624-E10	A0A1N6MBZ0	SWISSPROT	Candida galli	0.3133389	3.032	0.516	4.148
S624-F07	BFJ94474	GENESEQP	Metschnikowia sp	0.3148168	2.912	0.563	4.742
S624-F08	C4B4V9	SWISSPROT	Corynebacterium glutamicum (Brevibacterium saccharolyticum)	0.3466282	3.068	0.553	4.694
S624-F10	A0A1N6MBZ0	SWISSPROT	Candida galli	0.3296145	3.2	0.53	4.28
S624-G07	BFJ94474	GENESEQP	Metschnikowia sp	0.3518836	3.236	0.551	4.466
S624-G08	C4B4V9	SWISSPROT	Corynebacterium glutamicum (Brevibacterium saccharolyticum)	0.3078721	3.02	0.545	4.658
S624-G10	A0A1N6MBZ0	SWISSPROT	Candida galli	0.3440856	3.02	0.529	4.166
S624-H07	BFJ94474	GENESEQP	Metschnikowia sp	0.31948	2.84	0.517	4.49
S624-H08	C4B4V9	SWISSPROT	Corynebacterium glutamicum (Brevibacterium saccharolyticum)	0.3131478	3.08	0.523	4.79
S624-H09	C4B4V9	SWISSPROT	Corynebacterium glutamicum (Brevibacterium saccharolyticum)	0.3360712	3.152	0.534	4.64
S624-H12	A0A1Y6JY60	SWISSPROT	Lactobacillus zymae	0.1811037	2.312	0.541	4.22
S625-A02	CAG57753	GENPEPT	Candida glabrata	0.2905378	3.224	0.398	4.748
S625-A07	A0A202G714	SWISSPROT	Clavispora lusitaniae (Candida lusitaniae)	0.3753871	3.26	0.515	4.772
S625-A09	CAG87483	GENPEPT	Debaryomyces hansenii	0.3616796	2.696	0.551	4.712
S625-B02	CAG57753	GENPEPT	Candida glabrata	0.2782116	3.248	0.418	4.778
S625-B04	CAG60202	GENPEPT	Candida glabrata	0.3096715	3.116	0.53	4.79

S625-C03	CAG58441	GENPEPT	Candida glabrata	0.3371924	3.56	0.433	4.796
S625-C06	A0A202G702	SWISSPROT	Clavispora lusitaniae (Candida lusitaniae)	0.3008664	3.02	0.543	4.736
S625-C09	CAG87483	GENPEPT	Debaryomyces hansenii	0.3595092	2.732	0.562	4.796
S625-C10	D7KH13	SWISSPROT	Arabidopsis lyrata subsp. Lyrata	0.4508352	4.292	0.513	4.55
S625-C11	EFP12DWXL	AHGP	Spathaspora passalidarum	0.4111762	3.812	0.495	4.808
S625-D05	A0A202G6Z7	SWISSPROT	Clavispora lusitaniae (Candida lusitaniae)	0.298149	2.876	0.533	4.736
S625-D06	A0A202G702	SWISSPROT	Clavispora lusitaniae (Candida lusitaniae)	0.3208814	3.452	0.519	4.736
S625-D07	A0A202G714	SWISSPROT	Clavispora lusitaniae (Candida lusitaniae)	0.3050702	2.864	0.569	4.628
S625-E01	C8TEF4	TREMBL	Candida arabinofermentans	0.3198353	3.236	0.478	4.736
S625-E07	A0A202G714	SWISSPROT	Clavispora lusitaniae (Candida lusitaniae)	0.3019518	3.02	0.563	4.808
S625-E09	CAG87483	GENPEPT	Debaryomyces hansenii	0.3138275	3.14	0.55	4.79
S625-F02	CAG57753	GENPEPT	Candida glabrata	0.2179967	2.852	0.49	4.748
S625-F04	CAG60202	GENPEPT	Candida glabrata	0.3717669	3.464	0	4.742
S625-F10	D7KH13	SWISSPROT	Arabidopsis lyrata subsp. Lyrata	0.4410499	4.328	0.585	4.7
S625-G02	CAG57753	GENPEPT	Candida glabrata	0.2625253	3.056	0.523	4.772
S625-G03	CAG58441	GENPEPT	Candida glabrata	0.348596	3.8	0.539	4.79
S625-G04	CAG60202	GENPEPT	Candida glabrata	-0.0053626	0.056	0.566	0
S625-G06	A0A202G702	SWISSPROT	Clavispora lusitaniae (Candida lusitaniae)	0.3377488	3.008	0.526	4.694
S625-G09	CAG87483	GENPEPT	Debaryomyces hansenii	0.3629195	2.852	0.521	4.766
S625-G10	D7KH13	SWISSPROT	Arabidopsis lyrata subsp. Lyrata	0.4335384	4.34	0.58	4.76
S625-H02	CAG58441	GENPEPT	Candida glabrata	0.3719989	3.944	0.487	4.79
S625-H04	A0A202G6Z7	SWISSPROT	Clavispora lusitaniae	0.386876	3.272	0.559	4.718

			(Candida lusitaniae)				
S625-H06	A0A202G702	SWISSPROT	Clavispora lusitaniae (Candida lusitaniae)	0.3359639	3.596	0.52	4.724
S625-H09	D7KH13	SWISSPROT	Arabidopsis lyrata subsp. Lyrata	0.4388982	4.64	0.531	4.796
S625-H12	EFP14W5DC	AHGP	Clavispora lusitaniae	-0.0021492	0.044	0.554	0
S626-A01	EFP1D9FNG	AHGP	Spathaspora arborariae	0.3836988	3.908	0.511	5
S626-A02	EFP3FBKC6	AHGP	Kluyveromyces marxianus	0.3719001	3.908	0.566	4.73
S626-A05	C5DDE9	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.348118	2.936	0.478	4.712
S626-A06	EFPC7NHF4	NZGP	Phytate as P enrichment B	0.3309501	3.368	0.487	4.79
S626-A07	A0A0P1KWW5	SWISSPROT	Lachancea quebecensis	0.3830406	2.792	0.479	4.712
S626-A08	EFP5FS42L	AHGP	Candida sojae	0.3399488	2.804	0.559	4.418
S626-A10	EFP5N972X	AHGP	Priceomyces haplophilus	0.3577001	3.104	0.531	4.82
S626-A12	EFP5NR67S	AHGP	Candida carpophila	0.3441407	3.284	0.55	4.76
S626-B01	EFP1D9FNG	AHGP	Spathaspora arborariae	0.410607	3.944	0.507	4.79
S626-B02	EFP3FBKC6	AHGP	Kluyveromyces marxianus	0.2902997	2.972	0.532	4.718
S626-B05	C5DDE9	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.2664803	2.816	0.477	4.724
S626-B06	EFPC7NHF4	NZGP	Phytate as P enrichment B	0.3144973	2.756	0.49	4.784
S626-B08	EFP5FS42L	AHGP	Candida sojae	0.415525	3.38	0.508	4.748
S626-C02	EFP3FBKC6	AHGP	Kluyveromyces marxianus	0.2870368	2.804	0.516	4.724
S626-C03	EFP3TVZL9	NZGP	Ogataea methanolica	0.4422794	3.44	0.538	4.706
S626-C06	EFPC7NHF4	NZGP	Phytate as P enrichment B	0.2997432	2.84	0.507	4.7
S626-C07	EFP7FXWPL	AHGP	Yarrowia galli	0.2852328	2.768	0.53	4.748
S626-C08	A0A0P1KWW5	SWISSPROT	Lachancea quebecensis	0.3214082	2.468	0.522	4.742
S626-C10	EFP5N972X	AHGP	Priceomyces haplophilus	0.3224425	3.332	0.569	4.784
S626-D02	EFP3FBKC6	AHGP	Kluyveromyces marxianus	0.264995	2.876	0.53	4.778

S626-D05	C5DDE9	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.2699201	2.84	0.398	4.676
S626-D06	EFPC7NHF4	NZGP	Phytate as P enrichment B	0.3071855	2.876	0.433	4.604
S626-D07	A0A0P1KWW5	SWISSPROT	Lachancea quebecensis	0.3113258	3.044	0.489	4.772
S626-D08	A0A0P1KWW5	SWISSPROT	Lachancea quebecensis	0.3547152	2.828	0.526	4.808
S626-E02	EFP3FBKC6	AHGP	Kluyveromyces marxianus	0.3072901	2.948	0.529	4.754
S626-E03	EFP3TVZL9	NZGP	Ogataea methanolica	0.3311183	2.552	0.514	4.742
S626-E05	C5DDE9	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.3992743	3.224	0.418	4.748
S626-E06	EFPC7NHF4	NZGP	Phytate as P enrichment B	0.269026	2.888	0.503	4.802
S626-E07	EFP7FXWPL	AHGP	Yarrowia galli	0.3442749	3.068	0.563	4.808
S626-E08	A0A0P1KWW5	SWISSPROT	Lachancea quebecensis	0.2999467	2.972	0.533	4.808
S626-E09	EFP7FXWPL	AHGP	Yarrowia galli	0.5353578	5	0.52	5
S626-F02	EFP3FBKC6	AHGP	Kluyveromyces marxianus	0.282295	2.912	0.557	4.736
S626-F05	C5DDE9	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.2788902	2.948	0.488	4.772
S626-F06	EFPC7NHF4	NZGP	Phytate as P enrichment B	0.3581	2.84	0.54	4.718
S626-F07	EFP7FXWPL	AHGP	Yarrowia galli	0.3038494	2.816	0.535	4.808
S626-F08	A0A0P1KWW5	SWISSPROT	Lachancea quebecensis	0.3316429	2.996	0.517	4.76
S626-F09	EFP5FS42L	AHGP	Candida sojae	0.3504822	4.196	0.515	4.556
S626-F11	EFP5NNT0L	AHGP	Wickerhamia fluorescens	0.4123326	3.752	0.557	4.814
S626-F12	EFP5NR67S	AHGP	Candida carpophila	0.3161864	3.248	0.513	4.772
S626-G01	EFP1D9FNG	AHGP	Spathaspora arborariae	0.4445404	4.364	0.534	4.808
S626-G05	C5DDE9	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.3459753	3.116	0.49	4.808
S626-G06	EFPC7NHF4	NZGP	Phytate as P enrichment B	0.3096595	2.96	0.537	4.82
S626-G07	EFP7FXWPL	AHGP	Yarrowia galli	0.3110631	2.984	0	4.76
S626-G08	A0A0P1KWW5	SWISSPROT	Lachancea quebecensis	0.3436644	2.9	0.543	4.814

S626-G09	EFP5FS42L	AHGP	Candida sojae	0.6534965	5	0	4.808
S626-G10	EFP5N972X	AHGP	Priceomyces haplophilus	0.2798032	2.648	0.517	4.778
S626-G11	EFP5NNT0L	AHGP	Wickerhamia fluorescens	0.3974499	3.776	0.551	4.784
S626-G12	EFP5NR67S	AHGP	Candida carpophila	0.3413135	3.356	0.524	4.652
S626-H01	EFP1D9FNG	AHGP	Spathaspora arborariae	0.4248934	4.76	0.571	4.796
S626-H02	EFP3FBKC6	AHGP	Kluyveromyces marxianus	0.3332038	3.476	0.504	4.754
S626-H05	C5DDE9	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.3954762	3.26	0.523	4.73
S626-H06	EFPC7NHF4	NZGP	Phytate as P enrichment B	0.3592149	3.14	0.539	4.664
S626-H07	EFP7FXWPL	AHGP	Yarrowia galli	0.2963875	2.72	0.566	4.742
S626-H08	A0A0P1KWW5	SWISSPROT	Lachancea quebecensis	0.3222558	3.128	0.519	4.73
S627-A01	EFP6PD54N	AHGP	Wickerhamomyces anomalus NRRL Y-366-8	0.2389499	3.356	0.585	4.742
S627-A02	EFP7FXX0N	AHGP	Yarrowia alimentaria	0.3124194	2.768	0.553	4.79
S627-A03	EFP6RQ8JN	NZGP	Scheffersomyces stipitis	0.3901355	3.728	0.505	4.766
S627-A04	EFP5QXT3D	AHGP	Yarrowia deformans	0.2990597	2.984	0.529	4.772
S627-A05	EFP5NRP7F	AHGP	Candida carpophila	0.3206118	2.84	0.514	4.7
S627-A07	EFP7J7B0Q	AHGP	Ilyonectria destructans	0.4098684	4.28	0.512	4.736
S627-A08	EFP6BNQR8	AHGP	Lachancea cidri	0.3626695	3.44	0.505	4.688
S627-A09	EFP7HS9KT	AHGP	Sugiyamaella xylanicola	0.3140772	3.416	0.538	4.736
S627-A10	EFP6RN7NJ	NZGP	Schwanniomyces occidentalis	0.3673295	3.476	0.514	4.688
S627-A11	EFPN276J	AHGP	Kluyveromyces wickerhamii	0.4077458	4.592	0.424	4.562
S627-A12	EFP7WSC34	AHGP	Spathaspora boniae	0.3671354	4.448	0.552	5
S627-B01	EFP6PD54N	AHGP	Wickerhamomyces anomalus NRRL Y-366-8	0.4247049	3.68	0.58	5
S627-B03	EFP6RQ8JN	NZGP	Scheffersomyces stipitis	0.4255134	3.788	0.532	4.796
S627-B04	EFP5QXT3D	AHGP	Yarrowia deformans	0.3095839	3.14	0.557	4.748
S627-B05	EFP5NRP7F	AHGP	Candida carpophila	0.3536919	2.756	0.437	4.688

S627-B08	EFP6BNQR8	AHGP	Lachancea cidri	0.3144515	3.404	0.518	4.748
S627-B09	M5P6N0	SWISSPROT	Bacillus sonorensis	0.3710606	3.716	0.556	5
S627-C01	EFP6PD54N	AHGP	Wickerhamomyces anomalus NRRL Y-366-8	0.4256281	3.836	0.563	4.79
S627-C02	EFP7FXX0N	AHGP	Yarrowia alimentaria	0.3160127	2.876	0.523	4.808
S627-C03	EFP6RQ8JN	NZGP	Scheffersomyces stipitis	0.3294535	3.452	0.534	4.592
S627-C04	EFP5QXT3D	AHGP	Yarrowia deformans	0.328909	3.08	0.504	4.778
S627-C05	EFP5NRP7F	AHGP	Candida carpophila	0.3260983	3.176	0.527	4.73
S627-C06	EFP6T76PV	AHGP	Debaryomyces hansenii	0.2816585	3.116	0.539	4.724
S627-C07	EFP7J7B0Q	AHGP	Ilyonectria destructans	0.3392304	4.364	0.535	4.814
S627-C08	EFP6BNQR8	AHGP	Lachancea cidri	0.3348904	3.608	0.541	5
S627-C09	M5P6N0	SWISSPROT	Bacillus sonorensis	0.2851211	3.32	0.523	4.784
S627-D01	EFP6PD54N	AHGP	Wickerhamomyces anomalus NRRL Y-366-8	0.4056184	3.836	0.551	4.82
S627-D02	EFP7FXX0N	AHGP	Yarrowia alimentaria	0.335375	3.068	0.545	5
S627-D08	EFP6BNRRN	AHGP	Lachancea cidri	0.3724667	3.944	0.544	4.796
S627-D09	V4KEI8	SWISSPROT	Eutrema salsugineum (Sisymbrium salsugineum)	0.2510854	3.176	0.521	4.814
S627-D10	C5DHA8	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.3600077	3.908	0.496	4.784
S627-D11	EFPN276J	AHGP	Kluyveromyces wickerhamii	0.488197	4.568	0.526	4.748
S627-D12	EFP7WSC34	AHGP	Spathaspora boniae	0.4247819	4.628	0.538	4.826
S627-E01	EFP6PD54N	AHGP	Wickerhamomyces anomalus NRRL Y-366-8	0.3911058	3.824	0.517	4.742
S627-E02	EFP7FXX0N	AHGP	Yarrowia alimentaria	0.3400147	2.696	0.554	4.82
S627-E03	EFP6RQ8JN	NZGP	Scheffersomyces stipitis	0.3386662	3.368	0.566	4.802
S627-E06	EFP6T76PV	AHGP	Debaryomyces hansenii	0.3892144	3.284	0.482	4.772
S627-E07	EFP7J7B0Q	AHGP	Ilyonectria destructans	0.4318408	4.316	0.519	5
S627-E08	EFP6BNQR8	AHGP	Lachancea cidri	0.3349506	3.356	0.534	5
S627-E09	M5P6N0	SWISSPROT	Bacillus sonorensis	0.3059135	3.296	0.544	4.772

S627-E10	C5DHA8	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.2923483	3.692	0.507	4.664
S627-F01	EFP6PD54N	AHGP	Wickerhamomyces anomalus NRRL Y-366-8	0.3292878	3.476	0.495	4.802
S627-F02	EFP7FXX0N	AHGP	Yarrowia alimentaria	0.2954949	2.792	0.511	4.754
S627-F03	EFP6RQ8JN	NZGP	Scheffersomyces stipitis	0.3801304	3.548	0.532	4.784
S627-F04	EFP5QXT3D	AHGP	Yarrowia deformans	0.3117767	3.032	0.513	4.76
S627-F06	EFP6T76PV	AHGP	Debaryomyces hansenii	0.3125684	3.5	0.509	4.82
S627-F08	EFP6BNQR8	AHGP	Lachancea cidri	0.3268411	3.44	0.568	4.826
S627-F09	M5P6N0	SWISSPROT	Bacillus sonorensis	0.3383721	3.548	0.466	4.646
S627-F10	C5DHA8	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.3058737	3.308	0.526	4.736
S627-F11	EFPN276J	AHGP	Kluyveromyces wickerhamii	0.3829311	4.472	0.504	4.766
S627-F12	EFP7WSC34	AHGP	Spathaspora boniae	0.434199	4.952	0.499	4.796
S627-G01	EFP6PD54N	AHGP	Wickerhamomyces anomalus NRRL Y-366-8	0.4454809	3.992	0.451	4.772
S627-G03	EFP6RQ8JN	NZGP	Scheffersomyces stipitis	0.3035672	3.44	0.516	4.802
S627-G05	EFP5NRP7F	AHGP	Candida carpophila	0.3447082	2.756	0.53	4.73
S627-G07	EFP7J7B0Q	AHGP	Ilyonectria destructans	0.3399558	3.74	0.557	5
S627-G09	M5P6N0	SWISSPROT	Bacillus sonorensis	0.1799801	2.612	0.529	4.664
S627-G10	C5DHA8	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.3450805	3.728	0.512	4.718
S627-G11	EFPN276J	AHGP	Kluyveromyces wickerhamii	0.3135951	3.932	0.541	4.796
S627-G12	P49374	SWISSPROT	Kluyveromyces lactis (Candida sphaerica)	0.4695402	4.664	0.521	4.82
S627-H01	EFP6PD54N	AHGP	Wickerhamomyces anomalus NRRL Y-366-8	0.3672795	3.932	0.502	5
S627-H02	EFP7FXX0N	AHGP	Yarrowia alimentaria	0.3367283	3.092	0.55	4.658
S627-H04	EFP5QXT3D	AHGP	Yarrowia deformans	0.3051535	2.696	0.514	4.718

S627-H06	EFP6T76PV	AHGP	Debaryomyces hansenii	0.2812393	3.284	0.524	4.742
S627-H08	EFP6BNQR8	AHGP	Lachancea cidri	0.35157	3.512	0.547	4.82
S627-H09	M5P6N0	SWISSPROT	Bacillus sonorensis	0.2849132	3.14	0.548	4.742
S627-H10	C5DHA8	TREMBL	Lachancea thermotolerans (Kluyveromyces thermotolerans)	0.3143899	3.308	0.547	4.814
S627-H12	EFP7WSC34	AHGP	Spathaspora boniae	0.3941195	4.112	0.558	5
S628-A04	EFP7FXWQF	AHGP	Yarrowia galli	0.3549156	3.368	0.552	4.802
S628-A05	G8ZV29	TREMBL	Torulaspora delbrueckii (Candida colliculosa)	0.1746128	2.744	0.543	4.664
S628-A08	EFPB917WB	AHGP	Talaromyces adpressus	0.2170824	3.116	0.572	4.718
S628-A10	EFPBZD7P6	AHGP	Spathaspora sp.	0.3356947	3.884	0.57	4.814
S628-B02	EFP8ZW2C9	AHGP	Scheffersomyces stambukii	0.4062289	4.448	0.581	5
S628-C02	EFP8ZW2C9	AHGP	Scheffersomyces stambukii	0.3988605	4.04	0.564	5
S628-D02	EFP8ZW2C9	AHGP	Scheffersomyces stambukii	0.3913848	4.364	0.55	4.82
S628-D04	EFP7FXWQF	AHGP	Yarrowia galli	0.3393513	3.32	0.558	4.796
S628-D05	G8ZV29	TREMBL	Torulaspora delbrueckii (Candida colliculosa)	0.2987544	3.392	0.553	4.73
S628-D06	EFP6T73D9	AHGP	Debaryomyces hansenii	0.3294894	3.128	0.505	4.694
S628-D07	EFPB917WB	AHGP	Talaromyces adpressus	0.2596833	3.248	0.547	4.7
S628-D09	EFPBZD7P6	AHGP	Spathaspora sp.	0.3956318	4.052	0.552	5
S628-E03	EFP9CLGNG	NZGP	Meyerozyma caribbica	0.4031216	4.004	0.546	4.76
S628-E04	EFP7FXWQF	AHGP	Yarrowia galli	0.3794871	3.512	0.574	4.748
S628-E05	G8ZV29	TREMBL	Torulaspora delbrueckii (Candida colliculosa)	0.3243947	3.38	0.547	4.784
S628-E06	EFP6T73D9	AHGP	Debaryomyces hansenii	0.333262	3.908	0.586	4.796
S628-E07	EFPB917WB	AHGP	Talaromyces adpressus	0.2850768	3.368	0.546	4.694
S628-E08	EFPBZBQPC	AHGP	Metschnikowia fructicola	0.4053061	4.34	0.54	4.742
S628-E09	EFPBZD7P6	AHGP	Spathaspora sp.	0.397786	4.4	0.51	4.712
S628-F01	CAG57753	GENPEPT	Candida glabrata	0.3230515	4.208	0.551	4.73
S628-F05	G8ZV29	TREMBL	Torulaspora delbrueckii	0.3387605	3.356	0.513	4.658

			(Candida colliculosa)				
S628-F06	EFP6T73D9	AHGP	Debaryomyces hansenii	0.334762	3.356	0.545	4.784
S628-F07	EFPB917WB	AHGP	Talaromyces adpressus	0.3125936	3.512	0.538	4.718
S628-F08	EFPBZBQPC	AHGP	Metschnikowia fructicola	0.3727469	3.692	0.555	4.64
S628-F11	EFPL4075	AHGP	Kluyveromyces aestuarii	0.3582427	3.32	0.576	4.73
S628-G06	EFP6T73D9	AHGP	Debaryomyces hansenii	0.3060752	3.848	0.573	4.766
S628-G10	P39004	SWISSPROT	Saccharomyces cerevisiae	0.4065074	3.716	0.585	4.73
S628-H07	EFPB917WB	AHGP	Talaromyces adpressus	0.2842459	3.296	0.518	4.742
S628-H09	EFPBZD7P6	AHGP	Spathaspora sp.	0.4673099	4.532	0.542	4.808
S628-H11	EFP7FXWHQ	AHGP	Candida phangngaensis	0.3513347	3.272	0.548	4.778
S629-A02	EFP6BNRRN	AHGP	Lachancea cidri	0.3684927	4.328	0.524	5
S629-A03	V4KEI8	SWISSPROT	Eutrema salsugineum (Sisymbrium salsugineum)	0.3757136	4.316	0.541	5
S629-A04	A0A0P1KXV6	TREMBL	Lachancea quebecensis	0.3534477	3.572	0.538	4.736
S629-A06	EFP5QNR84	AHGP	Ambrosiozyma monospora	0.4085733	5	0.507	5
S629-A07	A0A1Q2ZT88	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.3123939	3.68	0.557	4.712
S629-A08	EFP5Q5H1L	AHGP	Ogataea methanolica	0.3691541	3.068	0.551	4.802
S629-A09	A0A1G4MC24	TREMBL	Lachancea fermentati (Zygosaccharomyces fermentati)	0.3087557	3.344	0.558	4.742
S629-A10	J4U468	TREMBL	Saccharomyces kudriavzevii	0.3881162	3.464	0.513	4.802
S629-B01	EFPL4075	AHGP	Kluyveromyces aestuarii	0.2675191	3.2	0.536	4.646
S629-B02	EFP6BNRRN	AHGP	Lachancea cidri	0.4160271	4.952	0.512	4.742
S629-B03	V4KEI8	SWISSPROT	Eutrema salsugineum (Sisymbrium salsugineum)	0.3826557	3.632	0.544	4.784
S629-B04	A0A0P1KXV6	TREMBL	Lachancea quebecensis	0.2628053	3.224	0.556	4.736
S629-B06	EFP5QNR84	AHGP	Ambrosiozyma monospora	0.4645669	5	0.526	4.712

S629-B07	A0A1Q2ZT88	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.2565069	3.152	0.552	4.472
S629-B08	EFP5Q5H1L	AHGP	Ogataea methanolica	0.3626641	3.116	0.581	5
S629-C01	R0GW82	SWISSPROT	Capsella rubella	0.3558343	4.784	0.56	4.766
S629-C04	A0A0P1KXV6	TREMBL	Lachancea quebecensis	0.3728305	3.332	0.523	4.79
S629-C05	EFP6RN7NJ	NZGP	Schwanniomyces occidentalis	0.3465646	3.56	0.514	5
S629-C06	EFP5QNR84	AHGP	Ambrosiozyma monospora	0.4770512	5	0.512	5
S629-C07	A0A1Q2ZT88	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.2838372	3.212	0.538	4.79
S629-C10	EFP7G7KNB	AHGP	Saccharomycopsis fibuligera	0.3184938	3.464	0.53	4.784
S629-D03	V4KEI8	SWISSPROT	Eutrema salsugineum (Sisymbrium salsugineum)	0.334514	3.548	0.534	4.718
S629-D06	EFP5QNR84	AHGP	Ambrosiozyma monospora	0.4964815	5	0.547	4.79
S629-D07	A0A1Q2ZT88	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.3254843	3.032	0.499	4.4
S629-D08	EFP5Q5H1L	AHGP	Ogataea methanolica	0.2463679	2.792	0.55	4.73
S629-D09	A0A1G4MC24	TREMBL	Lachancea fermentati (Zygosaccharomyces fermentati)	0.3463687	3.356	0.574	4.832
S629-D10	EFP7G7KNB	AHGP	Saccharomycopsis fibuligera	0.3283284	3.38	0.505	4.832
S629-E01	R0GW82	SWISSPROT	Capsella rubella	0.3990959	4.736	0.552	4.802
S629-E02	EFP6BNRRN	AHGP	Lachancea cidri	0.3003438	4.016	0.557	4.808
S629-E03	V4KEI8	SWISSPROT	Eutrema salsugineum (Sisymbrium salsugineum)	0.3731774	4.364	0.568	4.466
S629-E04	A0A0P1KXV6	TREMBL	Lachancea quebecensis	0.1828722	2.792	0.521	4.766
S629-E06	EFP5QNR84	AHGP	Ambrosiozyma monospora	0.4183418	5	0.424	5
S629-E07	A0A1Q2ZT88	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.336168	3.296	0.521	5
S629-E08	EFP5Q5H1L	AHGP	Ogataea methanolica	0.2575524	2.708	0.547	4.664
S629-E09	A0A1G4MC24	TREMBL	Lachancea fermentati	0.2415466	3.056	0.564	4.688

			(Zygosaccharomyces fermentati)				
S629-E10	EFP7G7KNB	AHGP	Saccharomycopsis fibuligera	0.3400576	3.44	0.586	4.802
S629-F01	R0GW82	SWISSPROT	Capsella rubella	0.1829044	3.128	0.482	4.694
S629-F03	V4KEI8	SWISSPROT	Eutrema salsugineum (Sisymbrium salsugineum)	0.4472193	4.112	0.551	4.712
S629-F06	EFP5QNR84	AHGP	Ambrosiozyma monospora	0.4344536	5	0.526	4.826
S629-F07	A0A1Q2ZT88	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.3915323	3.644	0.558	4.802
S629-F08	EFP5Q5H1L	AHGP	Ogataea methanolica	0.4105202	3.584	0.546	4.7
S629-F09	A0A1G4MC24	TREMBL	Lachancea fermentati (Zygosaccharomyces fermentati)	0.276406	3.104	0.543	4.718
S629-F10	EFP7G7KNB	AHGP	Saccharomycopsis fibuligera	0.3386973	3.368	0.545	4.73
S629-G02	EFP6BNRRN	AHGP	Lachancea cidri	0.416067	5	0.505	4.778
S629-G03	V4KEI8	SWISSPROT	Eutrema salsugineum (Sisymbrium salsugineum)	0.4167335	3.824	0.547	5
S629-G04	A0A0P1KXV6	TREMBL	Lachancea quebecensis	0.3373712	3.344	0.544	4.754
S629-G05	EFP6RN7NJ	NZGP	Schwanniomyces occidentalis	0.336727	3.464	0.496	4.742
S629-G06	EFP5QNR84	AHGP	Ambrosiozyma monospora	0.4207392	5	0.504	5
S629-G07	A0A1Q2ZT88	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.2210438	3.008	0.519	4.766
S629-G08	EFP5Q5H1L	AHGP	Ogataea methanolica	0.3466536	3.2	0.552	4.772
S629-G09	A0A1G4MC24	TREMBL	Lachancea fermentati (Zygosaccharomyces fermentati)	0.3040876	3.116	0.553	4.766
S629-G10	EFP7G7KNB	AHGP	Saccharomycopsis fibuligera	0.306382	3.284	0.573	4.73
S629-H01	R0GW82	SWISSPROT	Capsella rubella	0.3887375	5	0.509	4.478
S629-H02	EFP6BNRRN	AHGP	Lachancea cidri	0.4163048	4.748	0.518	4.736
S629-H04	A0A0P1KXV6	TREMBL	Lachancea quebecensis	0.3473167	3.356	0.466	4.574
S629-H06	EFP5QNR84	AHGP	Ambrosiozyma monospora	0.5132886	5	0.541	4.688

S629-H07	A0A1Q2ZT88	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.3773637	3.632	0.52	4.706
S629-H09	A0A1G4MC24	TREMBL	Lachancea fermentati (Zygosaccharomyces fermentati)	0.3183343	3.212	0.547	5
S629-H10	EFP7G7KNB	AHGP	Saccharomycopsis fibuligera	0.3286604	3.392	0.547	4.778
S642-A01	A0A0L9VMD5	SWISSPROT	Phaseolus angularis (Vigna angularis)	0.3623247	3.584	0.546	4.832
S642-A05	K0KRN7	SWISSPROT	Wickerhamomyces ciferrii (Pichia ciferrii)	0.3066395	3.248	0.559	4.808
S642-A07	P40885	SWISSPROT	Saccharomyces cerevisiae	0.3043919	3.152	0.512	5
S642-A08	A0A1U9X406	SWISSPROT	Pisum sativum	0.3215425	1.496	0.505	4.802
S642-A10	EFP9SKDNC	NZGP	Penicillium tularense	0.3568236	3.584	0.523	4.778
S642-B05	K0KRN7	SWISSPROT	Wickerhamomyces ciferrii (Pichia ciferrii)	0.3143539	3.308	0.576	4.802
S642-B08	A0A1U9X406	SWISSPROT	Pisum sativum	0.3785406	3.032	0.518	4.742
S642-B09	K9FYP3	SWISSPROT	Penicillium digitatum	0.4737424	5	0.568	5
S642-B11	EFP8DZ469	AHGP	Torulaspora microellipsoides	0.3729545	3.728	0.514	5
S642-C01	BAO92398	GENESEQP	Penicillium chrysogenum	0.3279053	4.208	0.522	4.742
S642-C03	EFP401CL1	NZGP	Penicillium vulpinum	0.3921512	4.424	0.542	5
S642-C04	A0A0A8KZI3	SWISSPROT	Kluyveromyces dobzhanskii	0.4418333	4.568	0.57	4.424
S642-C05	K0KRN7	SWISSPROT	Wickerhamomyces ciferrii (Pichia ciferrii)	0.3196028	3.68	0.542	4.826
S642-D01	A0A0L9VMD5	SWISSPROT	Phaseolus angularis (Vigna angularis)	0.3495611	3.356	0.518	4.76
S642-D02	BAO92398	GENESEQP	Penicillium chrysogenum	0.516172	5	0.54	4.652
S642-D05	K0KRN7	SWISSPROT	Wickerhamomyces ciferrii (Pichia ciferrii)	0.2879234	3.2	0.548	4.706
S642-D07	P40885	SWISSPROT	Saccharomyces cerevisiae	0.3588446	3.236	0.535	4.688
S642-D09	K9FYP3	SWISSPROT	Penicillium digitatum	0.2010501	3.26	0.551	4.724
S642-D10	EFP9SKDNC	NZGP	Penicillium tularense	0.3586858	3.848	0.521	4.778

S642-D11	EFP8DZ469	AHGP	Torulaspora microellipsoides	0.3514168	3.5	0.496	4.79
S642-E02	BAO92398	GENESEQP	Penicillium chrysogenum	0.6238107	5	0.555	5
S642-E03	EFP401CL1	NZGP	Penicillium vulpinum	0.3650241	3.908	0.51	5
S642-E07	P40885	SWISSPROT	Saccharomyces cerevisiae	0.3760686	3.26	0.519	4.82
S642-E08	A0A1U9X406	SWISSPROT	Pisum sativum	0.3079204	2.984	0.544	4.784
S642-E09	K9FYP3	SWISSPROT	Penicillium digitatum	0.587445	5	0.547	4.808
S642-E11	EFP8DZ469	AHGP	Torulaspora microellipsoides	0.2722842	3.164	0.507	5
S642-F07	P40885	SWISSPROT	Saccharomyces cerevisiae	0.4095371	3.476	0.557	4.796
S642-F11	EFP8DZ469	AHGP	Torulaspora microellipsoides	0.3929808	3.728	0.526	4.802
S642-G03	EFPL4075	AHGP	Kluyveromyces aestuarii	0.3496731	3.38	0.504	4.826
S642-G05	K0KRN7	SWISSPROT	Wickerhamomyces ciferrii (Pichia ciferrii)	0.2993671	3.356	0.536	4.652
S642-G09	K9FYP3	SWISSPROT	Penicillium digitatum	0.3605162	5	0.538	4.784
S642-G10	EFP9SKDNC	NZGP	Penicillium tularense	0.3324987	3.62	0.466	4.808
S642-G11	EFP8DZ469	AHGP	Torulaspora microellipsoides	0.3772988	3.8	0.512	4.79
S642-H03	EFP401CL1	NZGP	Penicillium vulpinum	0.3964691	3.716	0.538	4.676
S642-H04	K0KRN7	SWISSPROT	Wickerhamomyces ciferrii (Pichia ciferrii)	0.3016011	3.248	0.51	4.814
S642-H05	K0KRN7	SWISSPROT	Wickerhamomyces ciferrii (Pichia ciferrii)	0.3958623	3.62	0.56	4.814
S642-H06	A0A0N7MLX3	TREMBL	Lachancea quebecensis	0.3310281	3.188	0.524	5
S642-H07	P40885	SWISSPROT	Saccharomyces cerevisiae	0.2783827	3.332	0.525	5
S642-H09	K9FYP3	SWISSPROT	Penicillium digitatum	0.3964429	5	0.556	4.682
S643-A06	Q6CG69	SWISSPROT	Yarrowia lipolytica (Candida lipolytica)	0.2407538	2.324	0.558	4.826
S643-A07	AAT95983	GENPEPT	Torulaspora delbrueckii	0.3054765	2.9	0.53	4.802
S643-A10	A0A371ENF9	TREMBL	Mucuna pruriens	0.2962377	2.816	0.504	4.796
S643-B04	A3DSX4	TREMBL	Phaseolus vulgaris	0.3158349	3.104	0.558	4.754
S643-B05	EFP1CHF3L5	NZGP	Penicillium brevicompactum	0.1393603	2.504	0.55	4.706

S643-B06	Q6CG69	SWISSPROT	Yarrowia lipolytica (Candida lipolytica)	0.3325211	3.044	0.574	5
S643-B07	AAT95983	GENPEPT	Torulaspora delbrueckii	0.3066138	2.816	0.505	4.796
S643-B08	EFP70FSPD	NZGP	Saccharomyces cerevisiae	0.2613837	3.056	0.538	4.772
S643-B11	EFPL4075	AHGP	Kluyveromyces aestuarii	0.3554258	3.656	0.51	5
S643-C03	P49374	SWISSPROT	Kluyveromyces lactis (Candida sphaerica)	0.3765046	3.896	0.499	4.79
S643-C04	A3DSX4	TREMBL	Phaseolus vulgaris	0.3087604	2.936	0.519	4.808
S643-C05	EFP1CHF3L5	NZGP	Penicillium brevicompactum	0.4645332	5	0.547	4.736
S643-C07	AAT95983	GENPEPT	Torulaspora delbrueckii	0.3329219	3.116	0.586	4.796
S643-C10	A0A371ENF9	TREMBL	Mucuna pruriens	0.3382588	2.96	0.538	4.808
S643-D11	EFPL4075	AHGP	Kluyveromyces aestuarii	0.3842494	3.572	0.559	4.82
S643-E04	A3DSX4	TREMBL	Phaseolus vulgaris	0.2564662	2.816	0.551	4.736
S643-E08	EFP70FSPD	NZGP	Saccharomyces cerevisiae	0.2712773	2.54	0.518	4.802
S643-E10	A0A371ENF9	TREMBL	Mucuna pruriens	0.3370063	3.008	0.57	4.808
S643-E11	EFPL4075	AHGP	Kluyveromyces aestuarii	0.3551094	3.368	0.576	4.73
S643-F02	A0A1S2Z5S7	SWISSPROT	Cicer arietinum	0.2724876	3.032	0.552	4.796
S643-F04	EFP8ZW2C9	AHGP	Scheffersomyces stambukii	0.3449525	3.14	0.581	4.808
S643-F06	Q6CG69	SWISSPROT	Yarrowia lipolytica (Candida lipolytica)	0.2598887	2.852	0.553	4.808
S643-F07	EFP70FSPD	NZGP	Saccharomyces cerevisiae	0.3208742	2.984	0.545	4.82
S643-F08	EFP70FSPD	NZGP	Saccharomyces cerevisiae	0.2769906	2.648	0.572	4.76
S643-G04	A3DSX4	TREMBL	Phaseolus vulgaris	0.3467262	3.152	0.564	4.748
S643-G07	AAT95983	GENPEPT	Torulaspora delbrueckii	0.4204522	4.232	0.573	4.814
S643-G08	EFP70FSPD	NZGP	Saccharomyces cerevisiae	0.3330312	2.804	0.57	4.7
S643-H06	Q6CG69	SWISSPROT	Yarrowia lipolytica (Candida lipolytica)	0.3295576	3.248	0.513	4.826
S643-H07	AAT95983	GENPEPT	Torulaspora delbrueckii	0.3776705	3.236	0.547	5
S643-H10	A0A371ENF9	TREMBL	Mucuna pruriens	0.2445174	1.856	0.585	4.724
S643-H11	EFPL4075	AHGP	Kluyveromyces aestuarii	0.3843023	3.656	0.536	4.82
S644-A01	P43581	SWISSPROT	Saccharomyces cerevisiae	0.3582082	3.296	0.56	5

S644-A02	EFP5QXVB6	AHGP	Yarrowia deformans	0.2941795	3.212	0.512	4.736
S644-A04	J5S3S1	TREMBL	Saccharomyces kudriavzevii	0.2922744	3.332	0.556	4.61
S644-A10	G3AF26	TREMBL	Spathaspora passalidarum	0.4132226	4.904	0.553	4.682
S644-A11	EFPBZZ6M7	AHGP	Zygosaccharomyces kombuchaensis	0.3777841	3.704	0.586	4.712
S644-B01	P43581	SWISSPROT	Saccharomyces cerevisiae	0.2359901	2.432	0.552	4.778
S644-B04	J5S3S1	TREMBL	Saccharomyces kudriavzevii	0.2557415	3.068	0.523	5
S644-B09	Q9LNV3	SWISSPROT	Arabidopsis thaliana	0.4382488	5	0.546	4.22
S644-C02	EFP5QXVB6	AHGP	Yarrowia deformans	0.2310544	2.912	0.519	5
S644-C03	EFP7HS9KT	AHGP	Sugiyamaella xylanicola	0.2684057	2.828	0.544	4.73
S644-C04	J5S3S1	TREMBL	Saccharomyces kudriavzevii	0.2918369	2.996	0.521	4.718
S644-C09	Q9LNV3	SWISSPROT	Arabidopsis thaliana	0.5601625	5	0.552	4.466
S644-C12	B2G4F7	SWISSPROT	Zygosaccharomyces rouxii (Candida mogii)	0.4005432	4.124	0.518	5
S644-D01	P43581	SWISSPROT	Saccharomyces cerevisiae	0.3160571	2.528	0.482	4.682
S644-D02	EFP5QXVB6	AHGP	Yarrowia deformans	0.3411362	3.26	0.557	5
S644-D09	Q9LNV3	SWISSPROT	Arabidopsis thaliana	0.3689273	5	0.558	4.604
S644-E04	J5S3S1	TREMBL	Saccharomyces kudriavzevii	0.2911174	2.66	0.544	4.742
S644-F04	J5S3S1	TREMBL	Saccharomyces kudriavzevii	0.2935457	3.164	0.466	5
S644-F09	Q9LNV3	SWISSPROT	Arabidopsis thaliana	0.5917417	5	0.564	4.394
S644-G02	EFP7HS9KT	AHGP	Sugiyamaella xylanicola	0.3253088	3.116	0.505	4.742
S644-G03	EFP7HS9KT	AHGP	Sugiyamaella xylanicola	0.4172602	3.152	0.547	4.82
S644-G10	G3AF26	TREMBL	Spathaspora passalidarum	0.5893462	5	0.53	4.772
S644-H02	EFP5QXVB6	AHGP	Yarrowia deformans	0.2715584	2.804	0.518	4.55
S644-H03	EFP7HS9KT	AHGP	Sugiyamaella xylanicola	0.3098324	3.092	0.538	4.634
S645-B07	K0KRN7	SWISSPROT	Wickerhamomyces ciferrii (Pichia ciferrii)	0.3790369	3.224	0.548	4.634
S645-D06	EFP7FXX0V	AHGP	Yarrowia alimentaria	0.2302682	2.96	0.576	1.502

S645-E06	EFP7FXX0V	AHGP	Yarrowia alimentaria	0.379333	3.752	0.542	1.67

This study shows that sugar transporter genes expressed in S509-CO4 background strains comprising an active pentose pathway increases overall arabinose uptake. Approximately 80% of strains evaluated in this campaign showed improved growth rate and arabinose consumption compared to parent strain S509-C04. The following sugar transporter genes show highest affect in arabinose consumption: A0A078DBU3 (SEQ ID NO: 111), A1C8W7 (SEQ ID NO: 131), B1H0U6 (SEQ ID NO: 123), B6HE12 (SEQ ID NO: 108), D7KH13 (SEQ ID NO: 99), EFP1D9FNG (SEQ ID NO: 63), EFP4VJQD (SEQ ID NO: 72), EFP5FS42L (SEQ ID NO: 53), EFP6BNRRN (SEQ ID NO: 40) and K9FYP3 (SEQ ID NO: 124).

In comparison to parent strain S509-C04, some strains showed improved levels in growth and xylose consumption. Some genes that contributed to arabinose consumption also may assist in the uptake of xylose. Since most strains showed close to complete consumption of xylose, the following genes are identified with strains that showed higher growth rate when grown in xylose: A5DPY9 (SEQ ID NO: 116), A0A1L0BZU1 (SEQ ID NO: 138) and A1DAC2 (SEQ ID NO: 97).

Example 3: Construction of yeast strains with FPS1 and GPD1 deletions, and expressing GapN

Deletion of FPS1 in yeast strain MEJI797

The *S. cerevisiae* strain MEJI797 (strain MBG5012 of WO2019/161227 further expressing the glucoamylase of SEQ ID NO: 229 and the alpha-amylase of SEQ ID NO: 130) was transformed with the annealed primers 1231855 and 1231856 (*infra*). To aid homologous recombination of the annealed primers at the FPS1 locus, a plasmid containing the nuclease MAD7 and a guide RNA specific to FPS1, pMLBA814 (Figure 4), was also used in the transformation. Annealed primers 1231855 and 1231856 and pMLBA814 were transformed into the *S. cerevisiae* strain MEJI797 following a yeast electroporation protocol. Transformants were selected on YPD+cloNAT to select for colonies that contain the MAD7 plasmid pMLBA814. Sixteen transformants were picked onto YPD (2% glucose) solid medium.

Deletion of FPS1 was verified by PCR with locus specific primers 1231853 and 1231854 (*infra*) which anneal outside of the coding sequence for FPS1. Of the 16 transformants tested, all tested strains were deleted for FPS1. One of these transformants was chosen and designated S840-A03. The *S. cerevisiae* strain S840-A03 was passaged in YPD (2% glucose) liquid culture to facilitate removal of the MAD7 plasmid pMLBA814. After growth in liquid culture, individual

CA 03158982 2022-5-19

colonies were isolated on YPD (2% glucose) solid medium. Sixteen individual colonies were picked onto YPD + cloNAT solid medium and YPD (2% glucose) solid medium. All cloNAT sensitive colonies were pooled together to make a culture of S840-A03 free of the MAD7 plasmid pMLBA814. S840-A03 is deleted for FPS1 in the *S. cerevisiae* strain MEJI797.

Construction of the X-3 5' homology and promoter HOR7 containing fragment (fragment 1)

Synthetic DNA containing 500 bps homology to the X-3 site and the *S. cerevisiae* promoter HOR7 was synthesized by Thermo Fisher Scientific and designated HP13. Primers 1230181 + 1230203 (*infra*) were used to amplify HP13 resulting in a 1200 bps linear DNA fragment used for transformation.

Construction of the NADP$^+$ dependent glyceraldehyde 3-phosphate dehydrogenase (GapN, Q59931) containing fragment (fragment 2)

Fragment 2 contained 50 bps 3' *S. cerevisiae* HOR7 promoter, the NADP$^+$ dependent glyceraldehyde 3-phosphate dehydrogenase (GapN, Q59931), the *S. cerevisiae* TEF1 terminator, and 500 bps homology to the X-3 site. The DNA sequence containing these features was PCR amplified from a yeast strain that had been previously engineered with these elements called S789-A06 using primers 1232910 and 1230928 (*infra*). This resulted in a 2000 bps linear DNA fragment used for transformation.

Integration of linear fragments to generate a yeast with heterologous expression of GapN (Q59931)

The *S. cerevisiae* strain S840-A03 was transformed with fragments 1 and 2 described above. To aid homologous recombination of the 2 fragments at the genomic site X-3 a plasmid containing the nuclease MAD7 and a guide RNA specific to X-3 (pMLBA647; Figure 5) was also used in the transformation. Fragments 1, 2, and pMLBA647 were transformed into the into the *S. cerevisiae* strain S840-A03 following a yeast electroporation protocol. Transformants were selected on YPD+cloNAT to select for colonies that contain the MAD7 plasmid pMLBA647. Eight transformants were picked onto YPD (2% glucose) solid medium. Integration of the GapN (Q59931) expression cassette at X-3 was verified by PCR with locus specific primers and subsequent sequence analysis. Primers 1218020 + 1230932 (*infra*) amplify a 3600 bps fragment if the GapN (Q59931) expression cassette is inserted at X-3, and would amplify a 1500 bps fragment for the wildtype locus. Four of the eight transformants picked contained the correct sequence at X-3 for integration of fragments 1 and 2. One of these transformants was kept and designated MLBA1040. The *S. cerevisiae* strain MLBA1040 was passaged in YPD (2% glucose) liquid culture to facilitate removal of the MAD7 plasmid pMLBA647. After growth in liquid culture,

CA 03158982 2022-5-19

individual colonies were isolated on YPD (2% glucose) solid medium. Sixteen individual colonies were picked onto YPD + cloNAT solid medium and YPD (2% glucose) solid medium. All cloNAT sensitive colonies were pooled together to make a culture of MLBA1040 free of the MAD7 plasmid pMLBA647. MLBA1040 is deleted for FPS1 and expresses GapN (Q59931) at X-3 in the *S. cerevisiae* strain MEJI797.

Deletion of GPD1 in yeast strain MLBA1040

The *S. cerevisiae* strain MLBA1040 was transformed with the annealed primers 1231388 and 1231389 (*infra*) to delete GPD1. To aid homologous recombination of the annealed primers at the GPD1 locus, a plasmid containing the nuclease MAD7 and a guide RNA specific to GPD1, pJDIN171 (Figure 6), was also used in the transformation. Annealed primers 1231388/1231389 (*infra*) and pJDIN171 were transformed into the *S. cerevisiae* strain MLBA1040 following a yeast electroporation protocol. Transformants were selected on YPD+cloNAT to select for colonies that contain the MAD7 plasmid pJDIN171. Eight transformants were picked onto YPD (2% glucose) solid medium. Deletion of GPD1 was verified by PCR with locus specific primers. Primers 1231386 and 1231387 anneal outside of the coding sequence for GPD1. All eight transformants were deleted for GPD1 as determined by PCR. One of these was kept and designated S859-C01, and was deleted for both GPD1 and FPS1 and expresses GapN (Q59931) at the X-3 locus in the *S. cerevisiae* strain MEJI797.

Table 8: Primers used in this example.

Primer	SEQ ID NO.	Sequence
1231388	469	TGTACACCCC CCCCCTCCAC AAACACAAAT ATTGATAATA TAAAGATTTA TTGGAGAAAG ATAACATATC ATACTTTCCC CCACTTTTTT
1231389	470	AAAAAAGTGG GGGAAAGTAT GATATGTTAT CTTTCTCCAA TAAATCTTTA TATTATCAAT ATTTGTGTTT GTGGAGGGGG GGGGTGTACA
1230181	471	AACGACAGCA CAAAGGAACT TTCAC
1230203	472	TTTTTATTAT TAGTCTTTTT TTTTTTTTGA CAATATCTGT ATGATTTG
1232910	473	ATCAAATCAT ACAGATATTG TCAAAAAAAA AAAAAGACTA ATAATAAAAA ATGACCAAGC AGTATAAGAA CTATGTAAAC GG
1230928	474	GGCTACTGAT TTTGTTAAGC AACTCATCAAG
1231853	475	AGATTGCCCG GCCCTTTTTG
1231854	476	AGGTGACCAG GCTGAGTTCA TG
1231855	477	TACCAAGTAC GCTCGAGGGT ACATTCTAAT GCATTAAAAG ACATGTGAGA AAGCAGGCAA GAAAAAGAAA CAAATAATAT AGACTGATAG

CA 03158982 2022-5-19

1231856	478	CTATCAGTCT ATATTATTTG TTTCTTTTTC TTGCCTGCTT TCTCACATGT CTTTTAATGC ATTAGAATGT ACCCTCGAGC GTACTTGGTA
1231386	479	CATTCCATTC ACATATCGTC TTTGGCC
1231387	480	CACATCTGAA ATCATCGTAA GGAACTTTG
1218020	481	GAGATGGCCT ATTGATATCA AG
1230932	482	GCATTCACCT AAATAGCTAC TGCTCTATTA ATACAG

Example 4: Construction of yeast strains with an active pentose fermentation pathway and expressing a heterologous sugar transporter in host S859-C01 (deleted for FPS1, GPD1, expressing GapN)

This example describes the construction of yeast cells expressing an active pentose fermentation pathway and heterologous sugar transporter. DNA fragments were designed to allow for homologous recombination into the XII-2 and INT1 locus of the yeast S859-C01. The resulting strains had the five gene pentose utilization pathway (XR, LAD, LXR, XDH, and XK) at the XII-2 locus and the sugar transporter at the INT1 locus.

Construction of the fragments used in construction of the strains

The linear DNA framents listed in Table 9 were generated by PCR amplification. For the fragments HP70, TH12, TH26, HP27 (Figures 7-10, respectively) the template DNA was plasmid DNA containing these fragments. For fragments PCR amplified from existing strains (S509-C04, S509-D11, S515-G04, and S622-F07 (see U.S. Provisional Application No. 63/024,010, filed May 13, 2020, the content of which is hereby incorporated by reference), genomic DNA was used as template DNA in the PCR amplification. The primers and template listed *supra* with sequence listed in Table 9 were used in a PCR reaction containing 5-50 ng of plasmid DNA as template, 0.1 mM each dATP, dGTP, dCTP, dTTP, 1X Phusion HF Buffer (Thermo Fisher Scienctific), and 2 units Phusion Hot Start DNA polymerase in a final volume of 50 µL. The PCR was performed in a T100™ Thermal Cycler (Bio-Rad Laboratories, Inc.) programmed for one cycle at 98°C for 30 seconds followed by 32 cycles each at 98°C for 10 seconds, 59°C for 20 seconds, and 72°C for 40 seconds with a final extension at 72°C for 10 minutes. Following thermocycling, the PCR reaction products gel isolated and cleaned up using the NucleoSpin Gel and PCR clean-up kit (Machery-Nagel).

Table 9. DNA used during transformations to generate S982-B01, S982-B03 and S982-D06 strains

DNA fragment description	Template (strain or plasmid)	5' primer	3' primer	Size (bp)	name
50bp pPGK1|G4N708tADH3| pTDH3|EFP6RPZ73|tPDC6	S509-C04	1230242	1230207	3262	HH66a

tPDC6|pADH1|A3GF74|tTEF1|pPMA1	S509-C04	1230179	1230199	4079	HH66b
pPMA1|Q70FD1|tSTE2| pTEF2|C5J3R8|tPRM9	S509-C04	1230171	1230205	4100	HH67
tPRM9|XII-2_3'	TH12	1230177	1230216	750	TH12
INT1_5'|pRPL18B	HP70	1230733	1230195	1200	HP70
50bp pRPL18B|A1C8W7|50bp tENO2	S622-F07	1233593	1233594	1645	HH68
tENO2|INT1_3'	TH26	1230176	1230736	685	TH26
50bp pPGK1|G4N708| tADH3|pTDH3| EFP6RPZ73| tPDC6|pADH1|A3GF74|tTEF1|pPMA1	S509-D11	1230242	1230199	7079	HH69
pPMA1|G3YG17|tSTE2|pTEF2|B6HI95|tPRM9	S509-D11	1230171	1230205	4149	HH70
50bp pPGK1|G4N708|tADH3|pTDH3| EFP6RPZ73|tPDC6	S515-G04	1230242	1230207	3262	HH71a
50bp pPGK1|G4N708|tADH3| pTDH3|EFP6RPZ73|tPDC6|pADH1| A3GF74|tTEF1|pPMA1	S515-G04	1230179	1230199	4079	HH71b
pPMA1|EFP2BNSMN|tSTE2| pTEF2|Q763T4|tPRM9	S515-G04	1230171	1230205	4140	HH72
XII-2 5'+pPGK1	HP27	1230183	1230191	1200	HP27

Integration of the DNA fragments containing the C5 sugar utilization pathway and C5 sugar transporter into strain S859-C01

The yeast S859-C01 was transformed with the DNA pieces indicted in Table 10. To aid homologous recombination of the DNA fragments at the genomic XII-2 and INT1 sites a plasmid containing MAD7 and guide RNA specific to XII-2 and INT1 (pMIBa771; Figure 11) was also used in the transformation. These components were transformed into the into S. cerevisiae strain S859-C01 following a yeast electroporation protocol. Transformants were selected on YPD+cloNAT to select for transformants that contain the Mad7 plasmid pMIBa771. Transformants were picked using a Q-pix Colony Picking System (Molecular Devices) to inoculate one well of 96-well plate containing YPD+cloNAT media. The plates were grown for 2 days then glycerol was added to 20% final concentration and the plates were stored at -80°C until needed. Integration of specific phospholipase construct was verified by PCR with locus specific primers and subsequent sequencing. The strains generated in this example are shown in Table 10.

Table 10. Final strains and DNA pieces used to generate the strains using CRISPR plasmids pMIBa771

Strain ID	Template strain	Pieces	Pentose Pathway Genes (XR, LAD, LXR, XDH, XK)	Transporter
S859-C01	-	-	None	None
S982-B01	S509-C04	HP27, HH66a, HH66b, HH67, TH12, HP70, HH68, TH26	G4N708, C5J3R8, Q70FD1, EFP6RPZ73, A3GF74	A1C8W7
S982-B03	S509-D11	HP27, HH69, HH70, TH12, HP70, HH68, TH26	G4N708, B6HI95, G3YG17, EFP6RPZ73, A3GF74	A1C8W7

S982-D06	S515-G04	HP27, HH71a, HH71b, HH72, TH12, HP70, HH68, TH26	G4N708, Q763T4, EFP2BNSMN, EFP6RPZ73, A3GF74	A1C8W7

Table 11. Primers used in this example.

Primer name	SEQ ID NO.	Primer sequence (5'-3')
1230171	483	ACTCAGCTTT GCTAAAGTG CAAAAAGTC
1230176	484	AGTGCTTTTA ACTAAGAAT TATTAGTCTT TTCTGC
1230177	485	ACAGAAGACG GGAGACACT AG
1230179	486	GCCATTAGTA GTGTACTCA AACGAATTAT TG
1230183	487	TCTTTTCGCG CCCTGGAAA
1230191	488	TGTTTTATAT TTGTTGTAAA AAGTAGATAA TTACTTCCTT GATG
1230195	489	TTTGTTTTTT GTTTTCTTCT AATTGATTTT TTCTTTCTAT TTCC
1230199	490	ATTGATATTG TTCGATAATT AAATCTTTCT TATCTTCTTA TTCTTTTC
1230205	491	ATTTTCAACA TCGTATTTTC CGAAGCGTTG
1230207	492	TTGACGTGGC TGAACAACAG TC
1230216	493	TCAGTCCAAT GACAGTATTT TCTCCTTCTC AC
1230242	494	ACAGATCATC AAGGAAGTAA TTATCTACTT TTTACAAC
1230733	495	AAAGAGGAAA CTTCAACGCT TCATTTGAAA ATC
1230736	496	AATTGTAGAA TACAAATACA TAAATAAGTG TGTTCCCGAA G
1233593	497	ACCAAAGGAA ATAGAAAGAA AAAATCAATT AGAAGAAAAC AAAAAACAAA ATGTATAGAA TTTCAAACAT CTATGTTCTA GCAG
1233594	498	ATGATGAAAA AATAAGCAGA AAAGACTAAT AATTCTTAGT TAAAAGCACT TTATACTACC TCAGCGTGTA CTGC

Example 5: Evaluation of yeast strains expressing a heterologous sugar transporter

This example describes the evaluation of yeast strains in corn mash fermentations, including the impact of five carbon sugar utilization gene expression on final ethanol titer in a corn mash fermentation supplemented with L-arabinose and D-xylose is studied. The yeast strains used in this example are listed in Table 12.

Table 12.

Strain ID	C5 Pathway Engineering (XR, LAD, LXR, XDH, XK)	C5 Transporter	Redox Balance Engineering
MeJi797	None	None	none
S859-C01	None	None	GAPN (Q59931), ΔGPD1, ΔFPS1
S982-B01	G4N708, C5J3R8, Q70FD1, EFP6RPZ73, A3GF74	A1C8W7	GAPN (Q59931), ΔGPD1, ΔFPS1

CA 03158982 2022-5-19

S982-B03	G4N708, B6HI95, G3YG17, EFP6RPZ73, A3GF74	A1C8W7	GAPN (Q59931), ΔGPD1, ΔFPS1
S982-D06	G4N708, Q763T4, EFP2BNSMN, EFP6RPZ73, A3GF74	A1C8W7	GAPN (Q59931), ΔGPD1, ΔFPS1

Corn mash fermentation procedure

Yeast strains were incubated overnight in 20 mL YPD media (6% w/v D-glucose, 2% peptone, 1% yeast extract) in 50 ml baffled shake flasks at 32°C at 150 rpm at 32°C. Cells were harvested after ~ 24 hours incubation. Cells were collected by centrifugation and washed in DI water prior to resuspending in 20 mls DI water for dosing. Industrially obtained liquefied corn mash, where liquefaction was carried out using the Fortiva product from Novozymes, was supplemented with 24 ppm Lactrol and 283 ppm of urea. This mash was supplemented with both L-arabinose and D-xylose (solid) to a target concentration of 10g/kg mash for each sugar. Simultaneous saccharification and fermentation (SSF) was performed via mini-scale fermentations. Approximately 5 g of C5 supplemented corn mash was added to 15 mL conical tubes. Each tube was dosed with 5 x 10^6 cells/g of mash with one of the yeast strains shown in Table 12 followed by the addition of 0.36 AGU/g of dry solids of an exogenous glucoamylase enzyme product (Innova Achieve F). Six replicate tube fermentations were conducted for each yeast strain. Glucoamylase and yeast dosages were administered based on the exact weight of corn slurry in each vial. Tubes were incubated at 32°C and mixed two to three times per day via brief vortex. After 70 hours fermentation time, tubes were centrifuged @3500 rpm for 5 min. Supernatant samples were filtered with 0.2 μm syringe filters into vials for analysis of final ethanol level via HPLC.

Results

Final ethanol level results are shown in Figure 12. Strains expressing genes for five carbon sugar utilization were able to achieve significantly higher final ethanol levels in C5 supplemented corn mash fermentations than the two control strains. The strain with the highest mean final ethanol (S982-D06) was over 2% higher than the S859-C01 parent and 4% higher than the MeJi797 reference control.

CA 03158982 2022-5-19

CLAIMS

Claim 1. A recombinant host cell comprising a heterologous polynucleotide encoding a sugar transporter, wherein the transporter has a mature polypeptide sequence with at least 60%, e.g., at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 257-397; and wherein the cell comprises an active pentose fermentation pathway.

Claim 2. The recombinant host cell of claim 1, wherein the cell comprises an active xylose fermentation pathway.

Claim 3. The recombinant host cell of claim 2, wherein the cell comprises one or more active xylose fermentation pathway genes selected from:
a heterologous polynucleotide encoding a xylose isomerase (XI), and
a heterologous polynucleotide encoding a xylulokinase (XK).

Claim 4. The recombinant host cell of claim 2 or 3, wherein the cell comprises one or more active xylose fermentation pathway genes selected from:
a heterologous polynucleotide encoding a xylose reductase (XR),
a heterologous polynucleotide encoding a xylitol dehydrogenase (XDH), and
a heterologous polynucleotide encoding a xylulokinase (XK).

Claim 5. The recombinant host cell of any one of claims 1-4, wherein the cell comprises an active arabinose fermentation pathway.

Claim 6. The recombinant host cell of claim 5, wherein the cell comprises one or more active arabinose fermentation pathway genes selected from:
a heterologous polynucleotide encoding a L-arabinose isomerase (AI),
a heterologous polynucleotide encoding a L-ribulokinase (RK), and
a heterologous polynucleotide encoding a L-ribulose-5-P4-epimerase (R5PE).

Claim 7. The recombinant host cell of claim 5 or 6, wherein the cell comprises one or more active arabinose fermentation pathway genes selected from:
a heterologous polynucleotide encoding an aldose reductase (AR),
a heterologous polynucleotide encoding a L-arabinitol 4-dehydrogenase (LAD),
a heterologous polynucleotide encoding a L-xylulose reductase (LXR),
a heterologous polynucleotide encoding a xylitol dehydrogenase (XDH) and

a heterologous polynucleotide encoding a xylulokinase (XK).

Claim 8. The recombinant host cell of any one of claims 1-7, wherein the cell further comprises a heterologous polynucleotide encoding a glucoamylase, alpha-amylase, phospholipase, trehalase, protease, or pullulanase.

Claim 9. The recombinant host cell of any one of claims 1-8, wherein the cell is capable of higher anaerobic growth rate on pentose (e.g., xylose and/or arabinose) compared to the same cell without the heterologous polynucleotide encoding a sugar transporter (e.g., under conditions described in Example 2).

Claim 10. The recombinant host cell of any one of claims 1-9, wherein the cell is capable of higher pentose (e.g., xylose and/or arabinose) consumption compared to the same cell without the heterologous polynucleotide encoding a sugar transporter at about or after 120 hours fermentation (e.g., under conditions described in Example 2).

Claim 11. The recombinant host cell of any one of claims 1-10, wherein the cell is capable of higher ethanol production compared to the same cell without the heterologous polynucleotide encoding a sugar transporter under the same conditions (e.g., after 40 hours of fermentation).

Claim 12. The recombinant host cell of any one of claims 1-11, wherein the cell further comprises a heterologous polynucleotide encoding a transketolase (TKL1) and/or a heterologous polynucleotide encoding a transaldolase (TAL1).

Claim 13. The recombinant host cell of any one of claims 1-12, wherein the cell further comprises a disruption to an endogenous gene encoding a glycerol 3-phosphate dehydrogenase (GPD) and/or a disruption to an endogenous gene encoding a glycerol 3-phosphatase (GPP).

Claim 14. The recombinant host cell of any one of claims 1-13, wherein the cell is a yeast cell (e.g., a *Saccharomyces cerevisiae* cell).

Claim 15. A composition comprising the recombinant host cell of any one of claims 1-14 and one or more naturally occurring and/or non-naturally occurring components, such as components are selected from the group consisting of: surfactants, emulsifiers, gums, swelling agents, and antioxidants.

CA 03158982 2022-5-19

Claim 16. A method of producing a derivative of a recombinant host cell of any one of claims 1-14, the method comprising:

(a) providing:

(i) a first host cell; and

(ii) a second host cell, wherein the second host cell is a recombinant host cell of any one of claims 1-14;

(b) culturing the first host cell and the second host cell under conditions which permit combining of DNA between the first and second host cells;

(c) screening or selecting for a derive host cell.

Claim 17. A method of producing a fermentation product from a starch-containing or cellulosic-containing material, the method comprising:

(a) saccharifying the starch-containing or cellulosic-containing material; and

(b) fermenting the saccharified material of step (a) with the recombinant host cell of any one of claims 1-14 under suitable conditions to produce the fermentation product.

Claim 18. The method of claim 17, wherein saccharification of step (a) occurs on a starch-containing material, and wherein the starch-containing material is either gelatinized or ungelatinized starch.

Claim 19. The method of claim 18, comprising liquefying the starch-containing material by contacting the material with an alpha-amylase prior to saccharification.

Claim 20. Use of a recombinant host cell of any one of claims 1-14 in the production of ethanol.

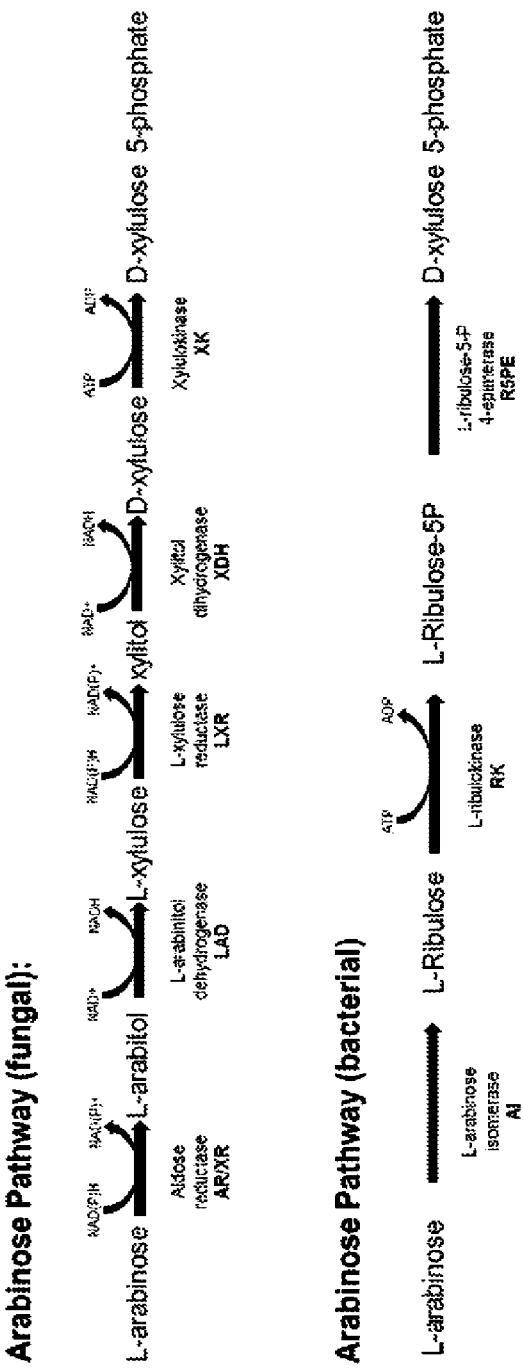


Figure 1

1

CA 03158982 2022-5-19

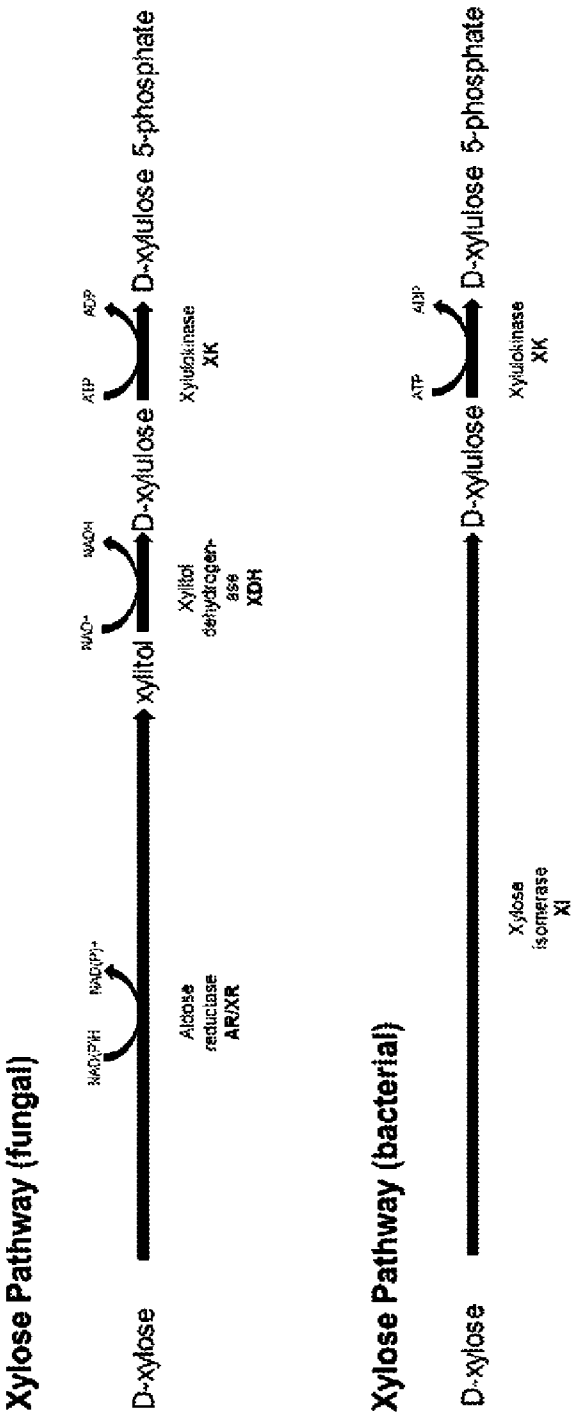


Figure 2

CA 03158982 2022-5-19

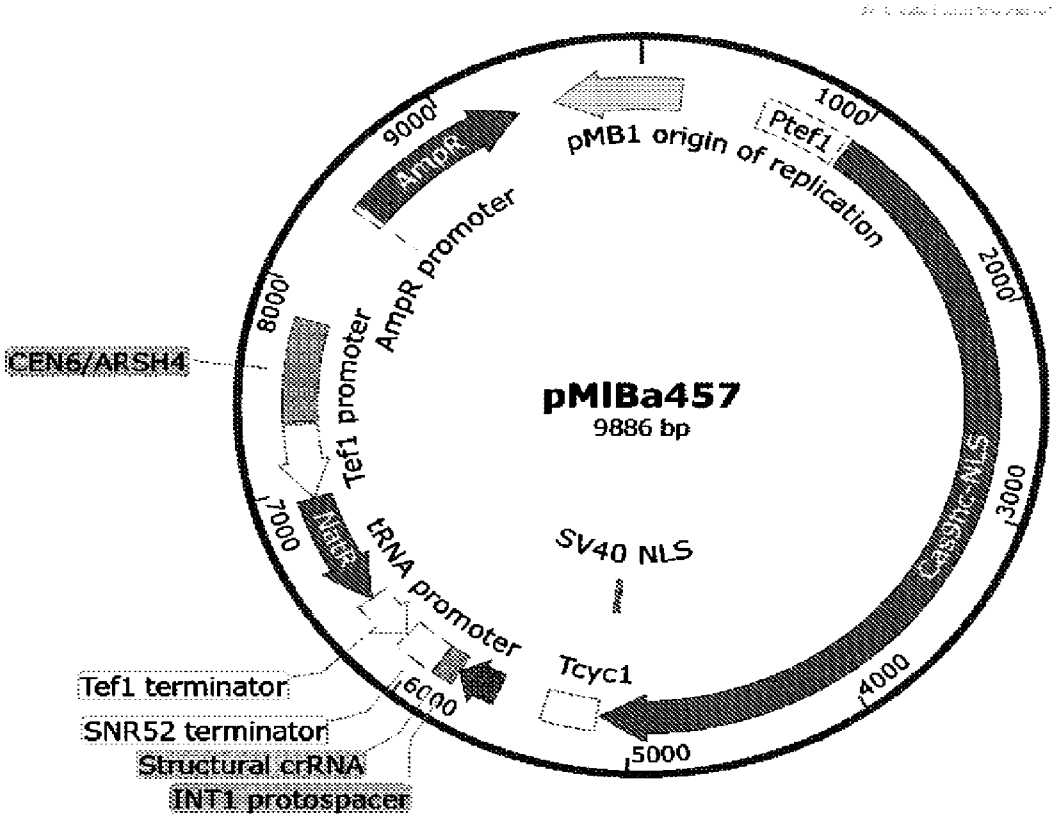


Figure 3

CA 03158982 2022-5-19

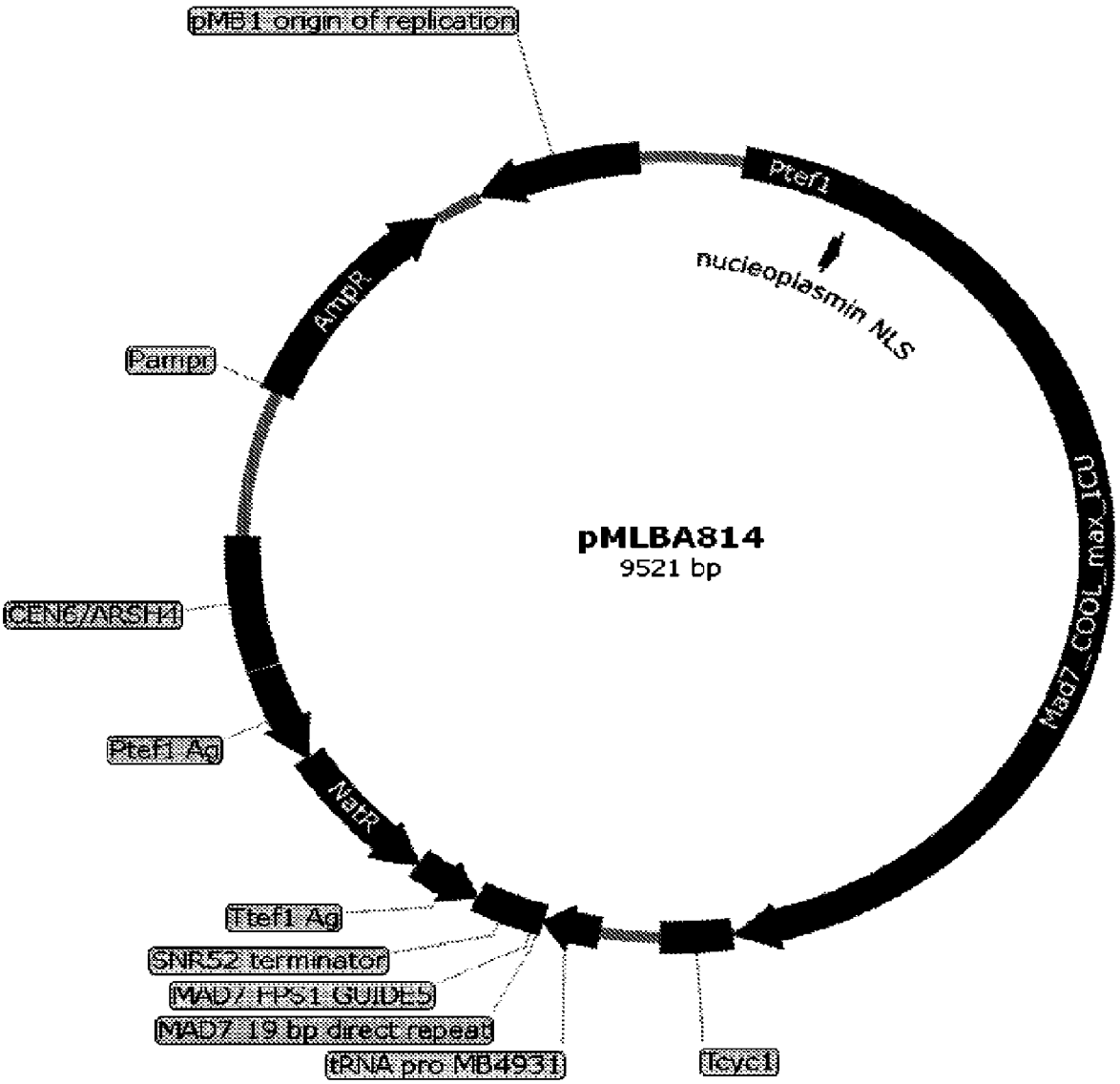


Figure 4

CA 03158982 2022-5-19

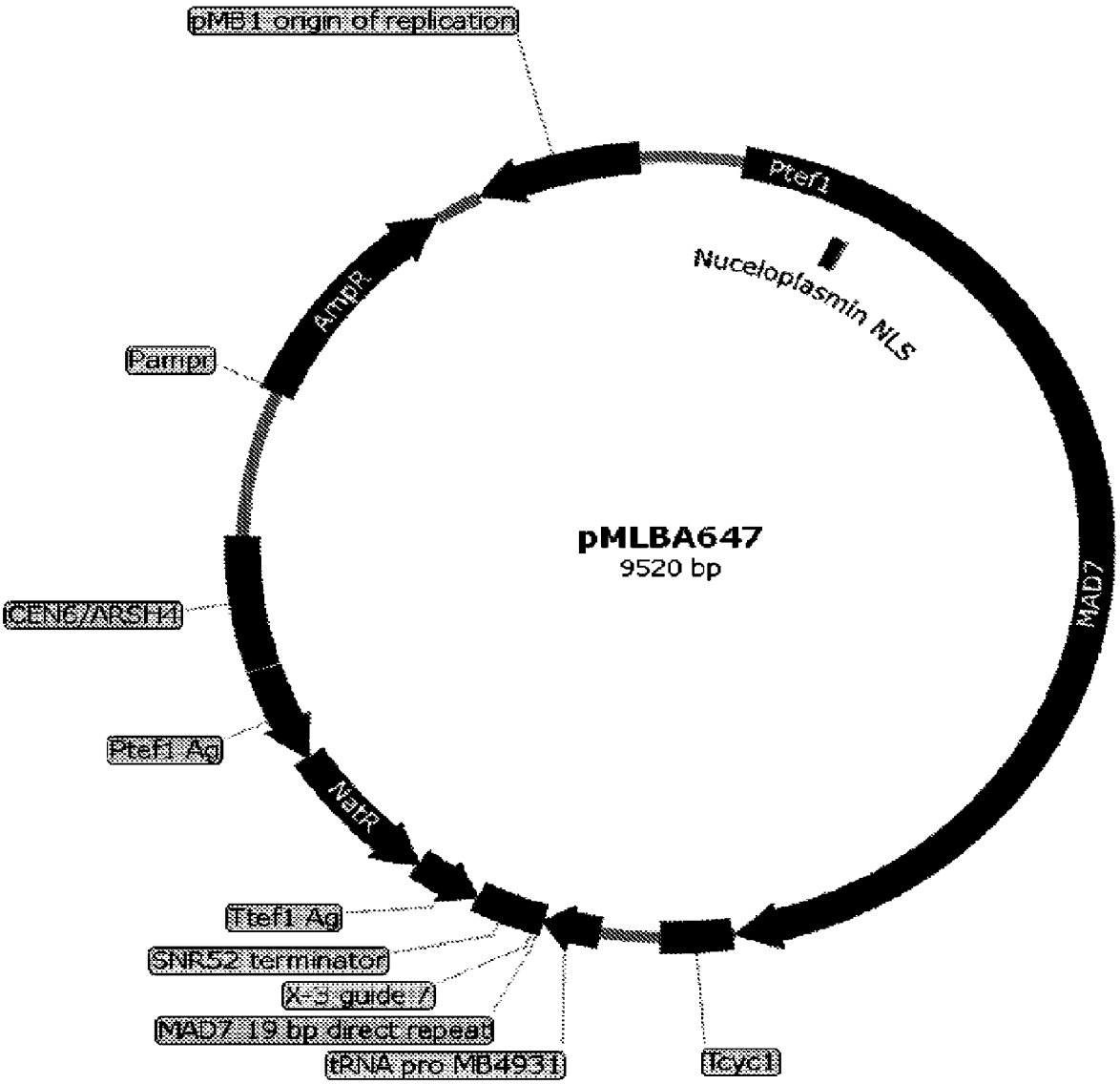


Figure 5

CA 03158982 2022-5-19

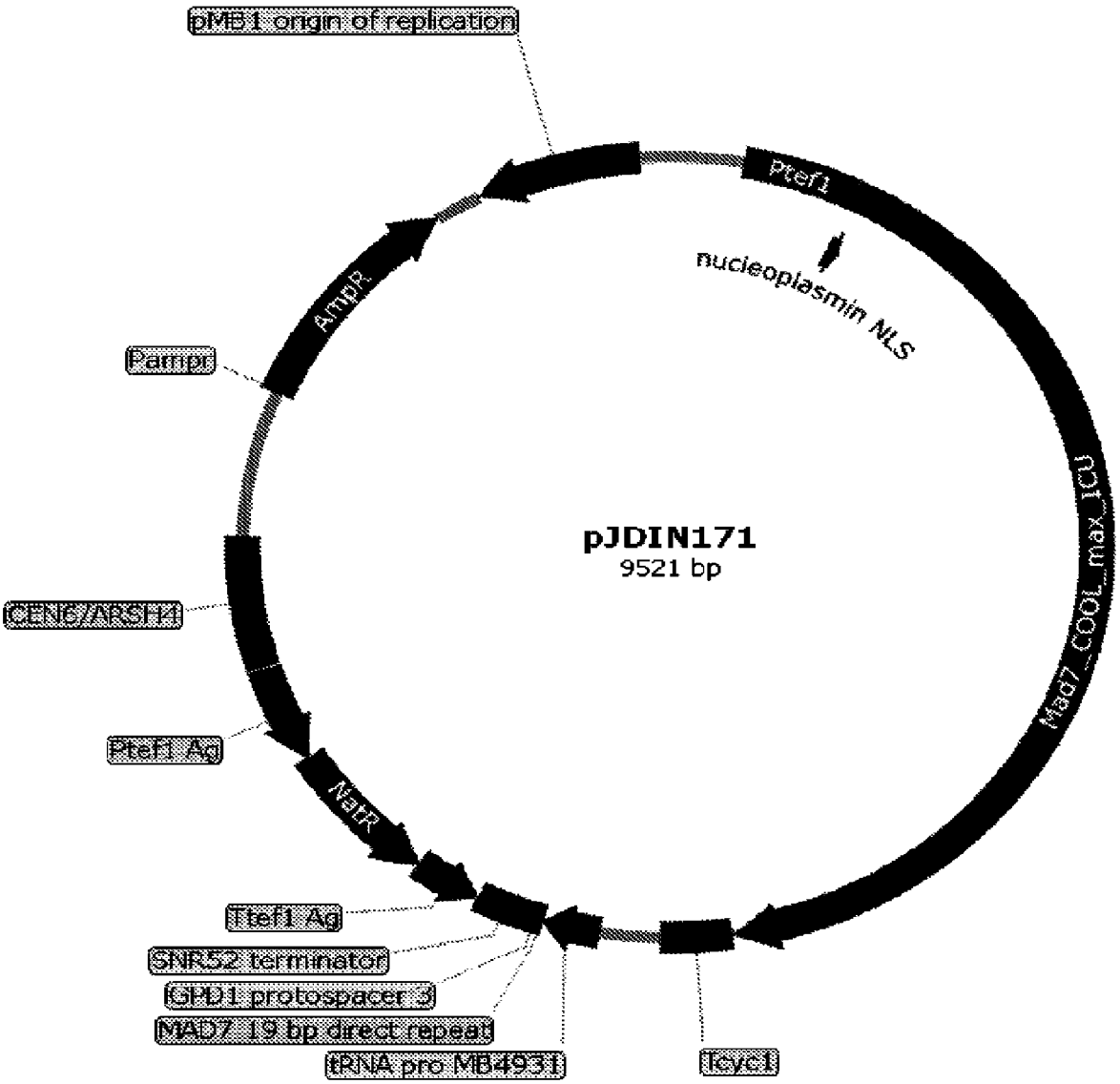


Figure 6

CA 03158982 2022-5-19

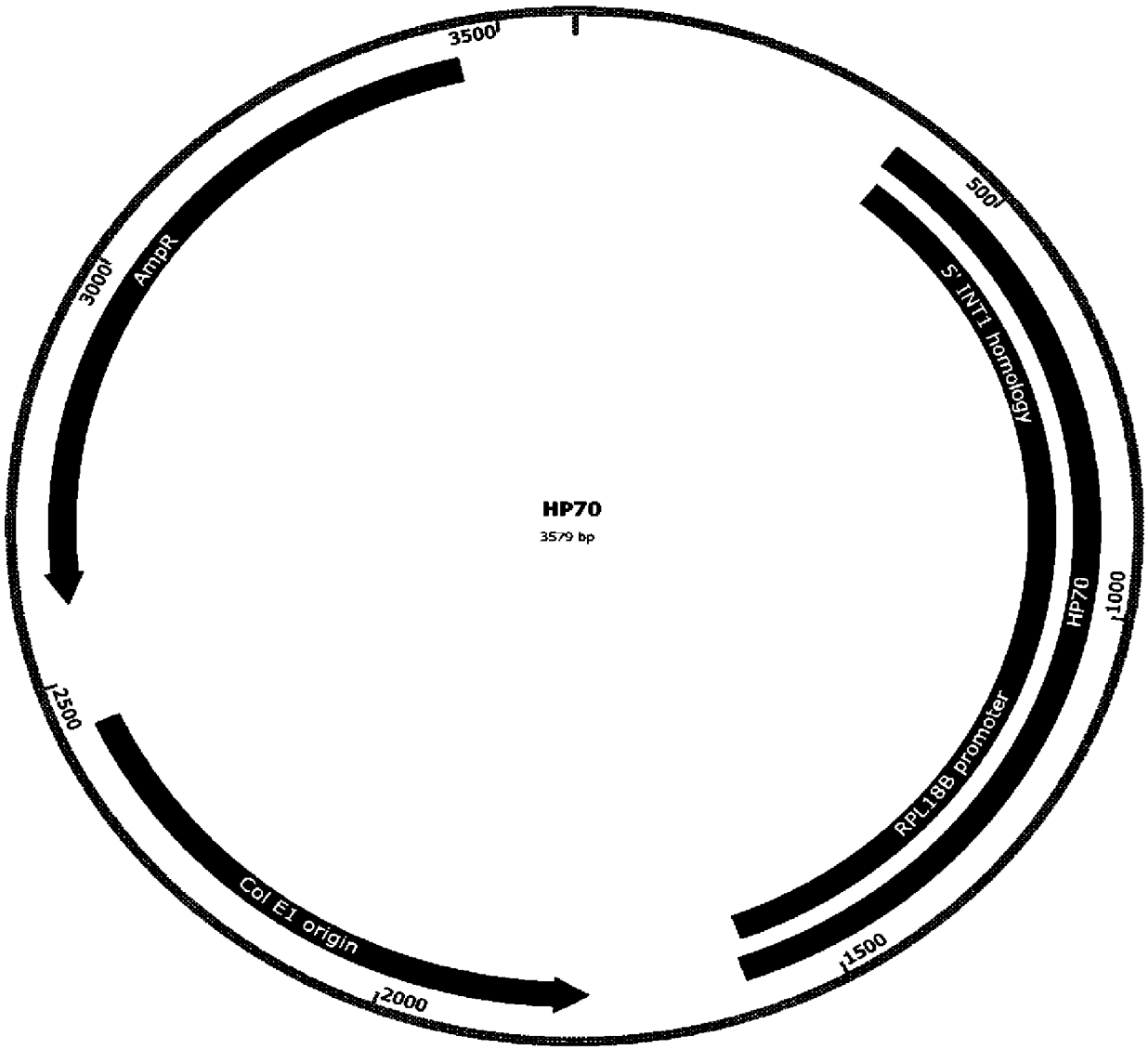


Figure 7

CA 03158982 2022-5-19

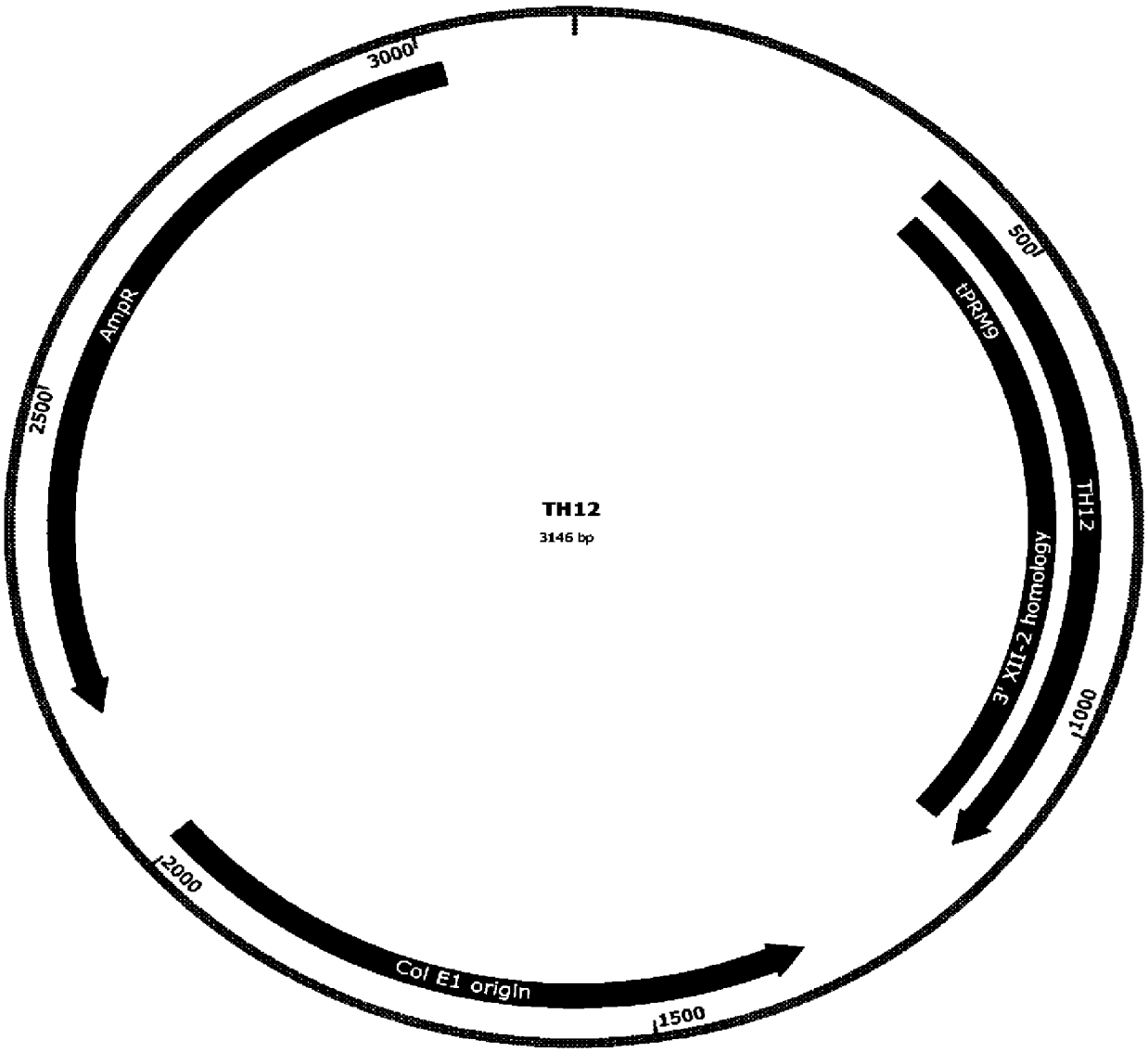


Figure 8

CA 03158982 2022-5-19

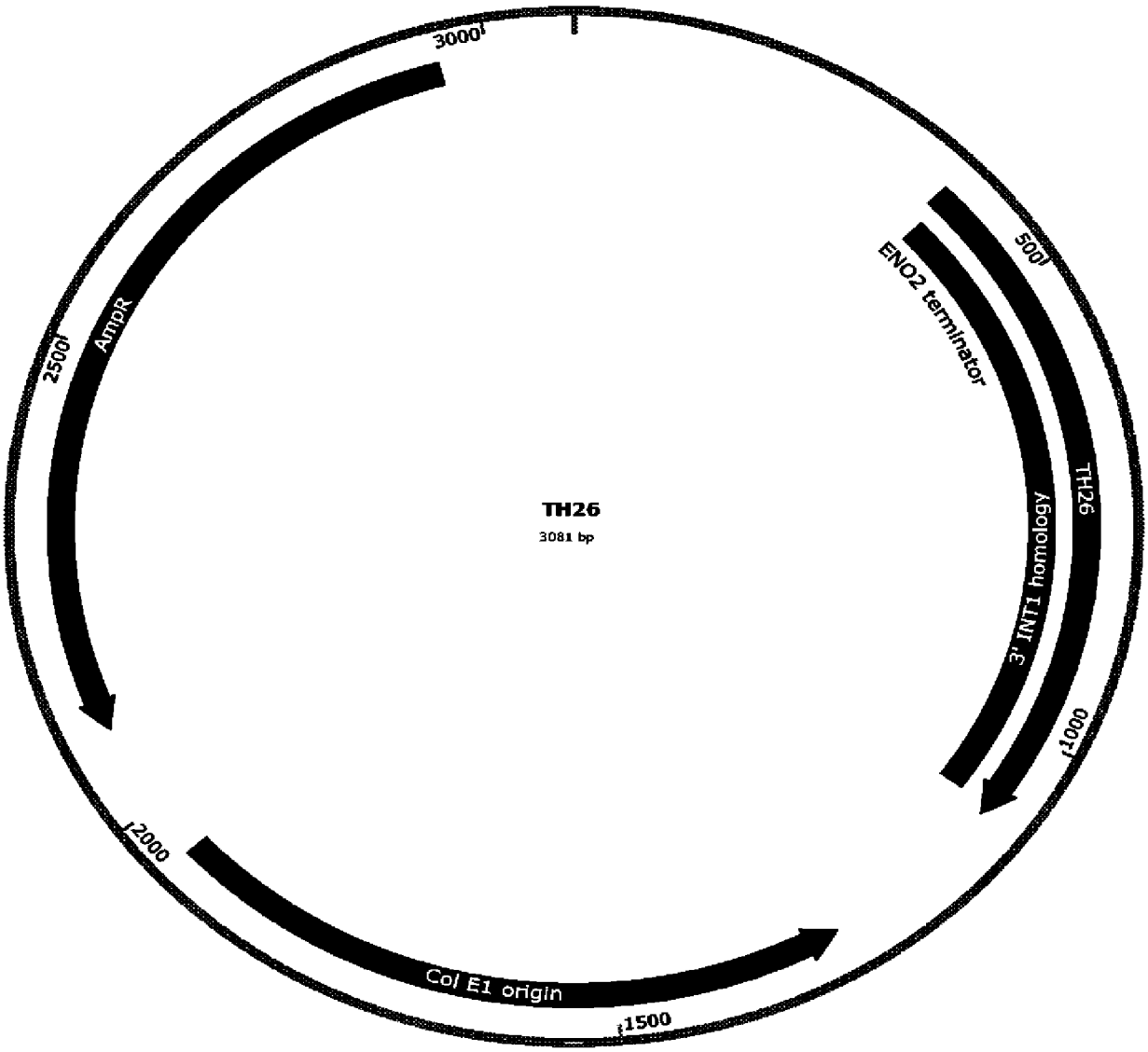


Figure 9

CA 03158982 2022-5-19

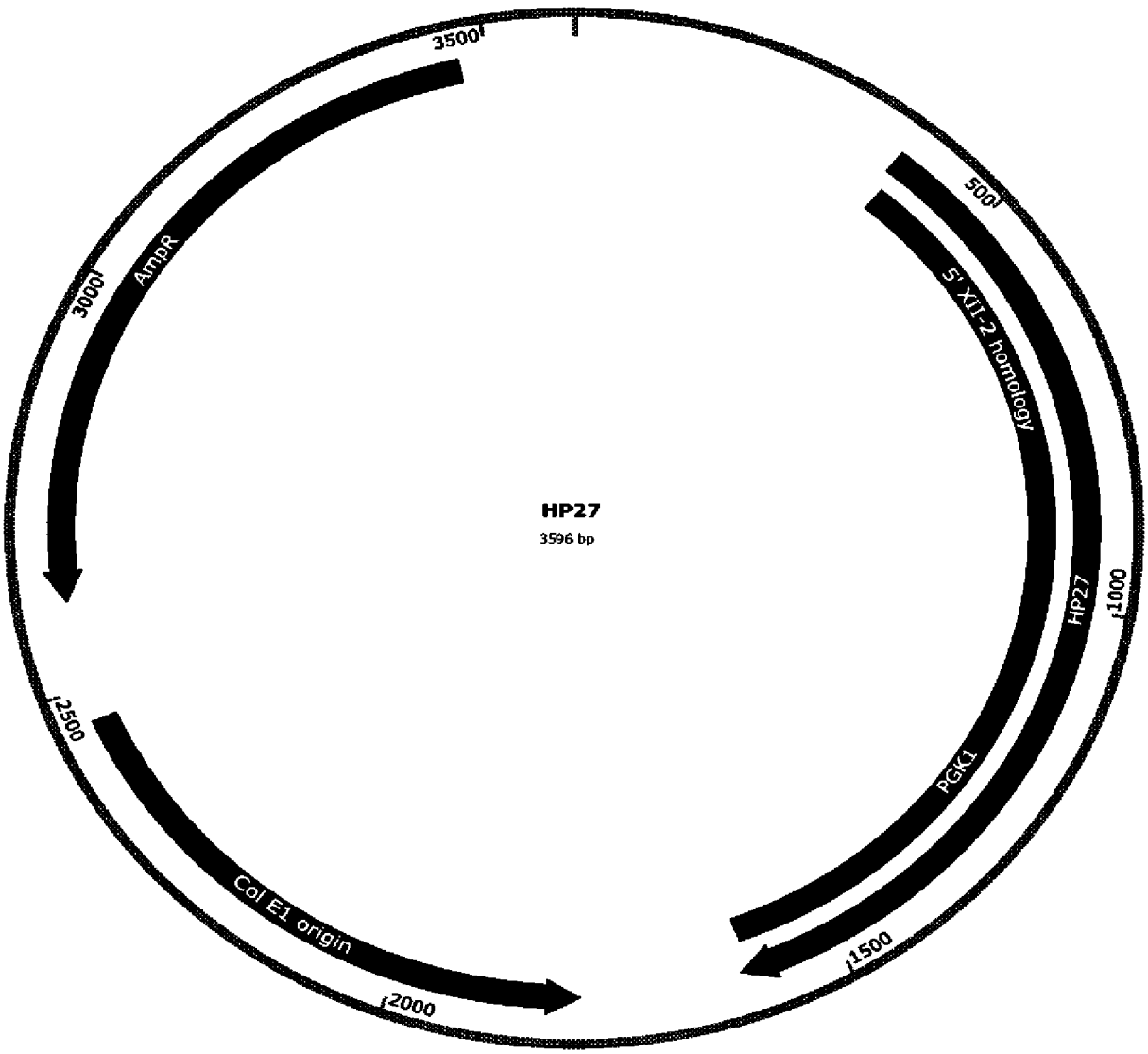


Figure 10

CA 03158982 2022-5-19

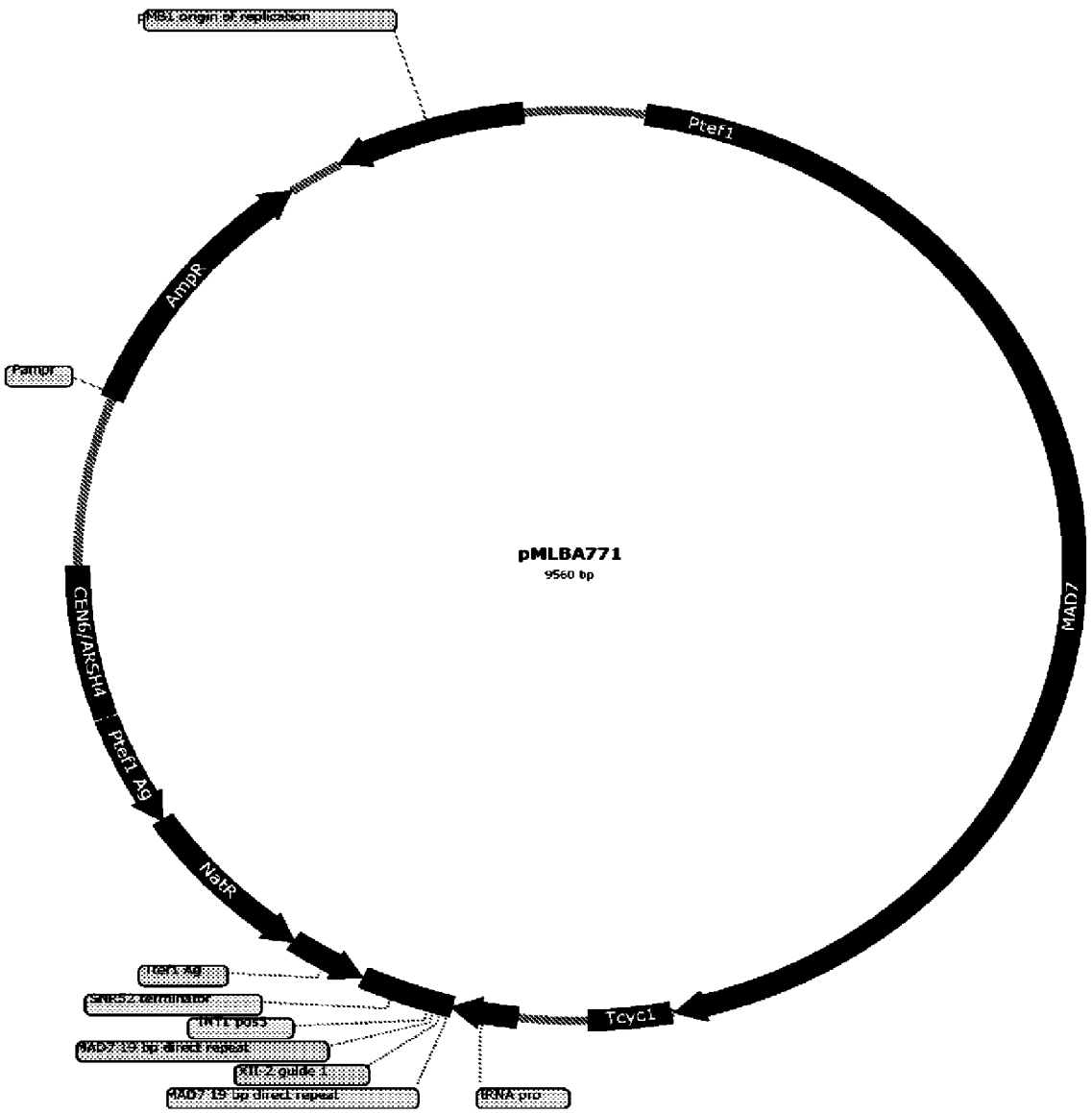


Figure 11

CA 03158982 2022-5-19

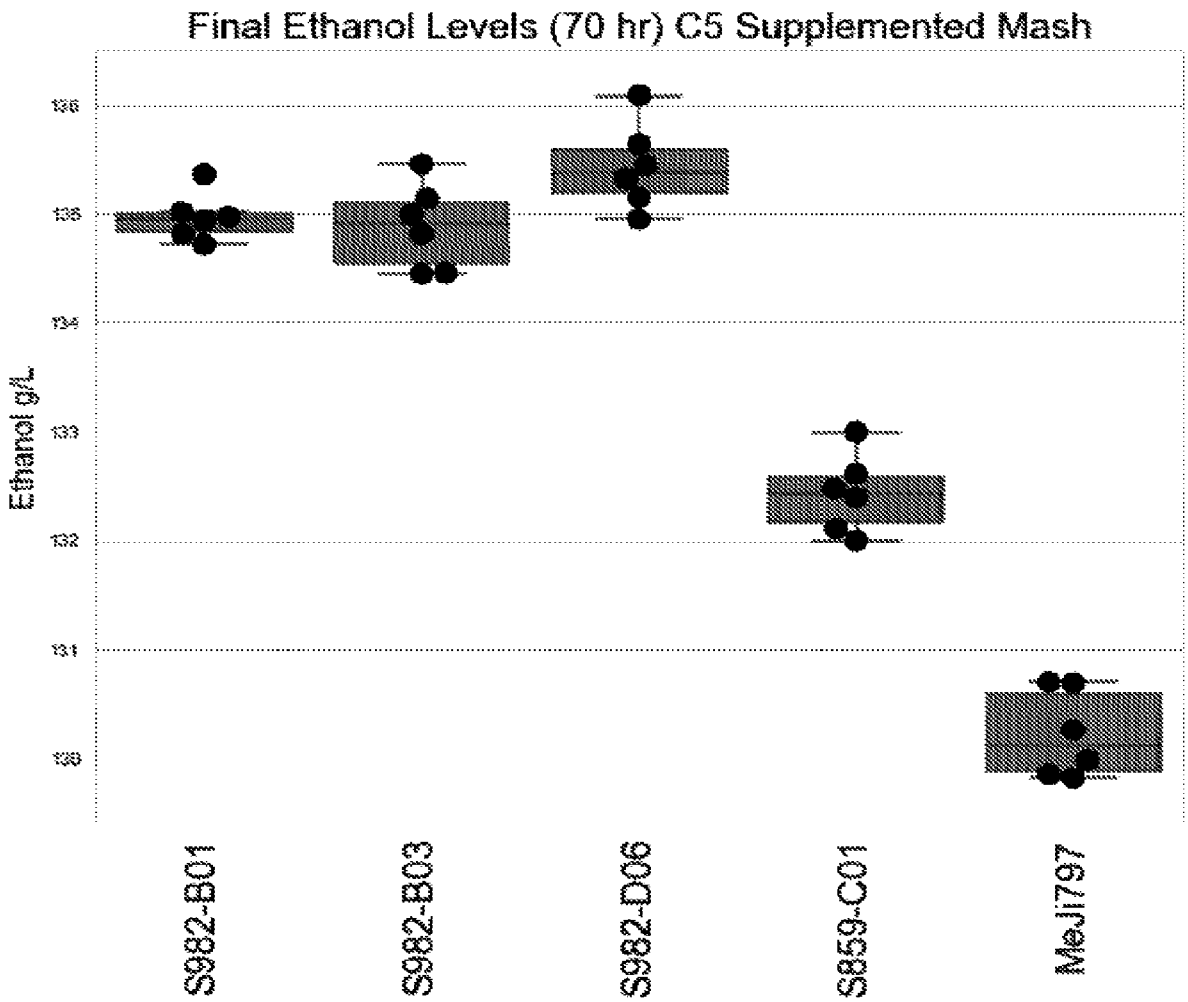


Figure 12

CA 03158982 2022-5-19