



(43) International Publication Date
18 September 2014 (18.09.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/140171 A1

(51) International Patent Classification:

C12P 19/02 (2006.01) C12N 9/42 (2006.01)
C12P 19/14 (2006.01) C12P 7/10 (2006.01)
C13K 1/02 (2006.01)

(21) International Application Number:

PCT/EP2014/054954

(22) International Filing Date:

13 March 2014 (13.03.2014)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/783,313 14 March 2013 (14.03.2013) US

(71) Applicants: **DSM IP ASSETS B.V.** [NL/NL]; Het Overloon 1, NL-6411 TE Heerlen (NL). **CONCORDIA UNIVERSITY** [—/CA]; 1455 de Maisonneuve Boulevard West, Suite GM 1000, Montréal, Québec H3G 1M8 (CA).

(72) Inventors: **TSANG, Adrian**; P.O. Box 4, NL-6100 AA Echt (NL). **BUTLER, Gregory**; P.O. Box 4, NL-6100 AA Echt (NL). **POWLOWSKI, Justin**; P.O. Box 4, NL-6100 AA Echt (NL).

(74) Agent: **MATULEWICZ, Emil Rudolf Antonius**; DSM Intellectual Property, P.O. Box 4, NL-6100 AA Echt (NL).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: NOVEL CELL WALL DECONSTRUCTION ENZYMES OF THIELAVIA AUSTRALIENSIS AND USES THEREOF

(57) Abstract: The present invention relates to a novel process for degrading biomass or pretreated biomass to sugars wherein an enzyme is used comprising a. a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 7- 9; b. a polypeptide comprising an amino acid sequence that is at least 85%, preferably at least 90%, more preferably at least 95%, even more preferably at least 96%, 97%, 98% or 99% amino acid sequence identity to the polypeptide defined in (a); c. a polypeptide comprising an amino acid sequence encoded by the nucleic acid sequence of any one of SEQ ID NOs: 1 -6; d. a polypeptide comprising an amino acid sequence encoded by a polynucleotide molecule that hybridizes under medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complement of a polynucleotide molecule comprising the nucleic acid sequence of any one of SEQ ID NOs: 1 -6; e. a polypeptide comprising an amino acid sequence encoded by a polynucleotide molecule having 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%, or at least 99% nucleic acid sequence identity to a polynucleotide comprising the nucleic acid sequence of any one of SEQ ID NOs: 1 -6; or f. a polypeptide comprising an amino acid sequence encoded by any one of the exonic nucleic acid sequences corresponding to positions: 1-92, 150-654, 815-1385, 1452-1714 of SEQ ID NO: 2; 1 -246, 324-1007, 1082-1 135 of SEQ ID NO: 1; 1-394, 506-789 of SEQ ID NO: 3.



WO 2014/140171 A1

NOVEL CELL WALL DECONSTRUCTION ENZYMES OF *THIELAVIA AUSTRALIENSIS*
AND USES THEREOF

5

FIELD OF THE INVENTION

The present invention relates to novel polypeptides and enzymes having activities relating to biomass processing and/or degradation (e.g., cell wall deconstruction), as well as polynucleotides, vectors, cells, compositions and tools relating to same, or functional variants thereof. More particularly, the present invention relates to secreted enzymes that may be isolated from the fungus, *Thielavia australiensis* strain ATCC 28236. Uses thereof in various industrial processes such as in biofuels, food preparation, animal feed, pulp and paper, textiles, detergents, waste treatment and others are also disclosed.

BACKGROUND OF THE INVENTION

Biomass-processing enzymes have a number of industrial applications such as in: the biofuel industry (e.g., improving ethanol yield and/or increasing the efficiency and economy of ethanol production); the food industry (e.g., production of cereal-based food products; the feed-enzyme industry (e.g., increasing the digestibility/absorption of nutrients); the pulp and paper industry (e.g., enhancing bleachability of pulp); the textile industry (e.g., treatment of cellulose-based fabrics); the waste treatment industry (e.g., de-colorization of synthetic dyes); the detergent industry (e.g., providing eco-friendly cleaning products); and the rubber industry (e.g., catalyzing the conversion of latex into foam rubber).

In particular, driven by the limited availability of fossil fuels, there is a growing interest in the biofuel industry for improving the conversion of biomass into second-generation biofuels. This process is heavily dependent on inexpensive and effective enzymes for the conversion of lignocellulose to ethanol. Cellulase enzyme cocktails involve the concerted action of endoglucanases, cellobiohydrolases, and beta-glucosidases. The current cost of cellulose-degrading enzymes is too high for bioethanol to compete economically with fossil fuels. Cost reduction may result from the discovery of cellulase enzymes with, for example, higher specific activity, lower production costs, and/or greater compatibility with processing conditions including temperature, pH and the presence of inhibitors in the biomass, or produced as the result of biomass pre-treatment.

Conversion of plant biomass to glucose may also be enhanced by supplementing cellulose cocktails with enzymes that degrade the other components of biomass, including hemicelluloses, pectins and lignins, and their linkages, thereby improving the accessibility of cellulose to the cellulase enzymes. Such enzymes include, without being limiting, to: xylanases, mannanases, arabinanases, esterases, glucuronidases, xyloglucanases and arabinofuranosidases for hemicelluloses; lignin peroxidases, manganese-dependent peroxidases, versatile peroxidases, and laccases for lignin; and pectate lyase, pectin lyase, polygalacturonase, pectin acetyl esterase, alpha-arabinofuranosidase, beta-galactosidase, galactanase, arabinanase, rhamnogalacturonase, rhamnogalacturonan lyase, and rhamnogalacturonan acetyl esterase, xylogalacturonosidase, xylogalacturonase, and rhamnogalacturonan lyase. Additionally, glycoside hydrolase family 61 (GH61) proteins have been shown to stimulate the activity of cellulase preparations.

These enzymes may also be useful for other purposes in processing biomass. For example, the lignin modifying enzymes may be used to alter the structure of lignin to produce novel materials, and hemicellulases may be employed to produce 5-carbon sugars from hemicelluloses, which may then be further converted to chemical products.

There is also a growing need for improved enzymes for food processing and feed applications. Cereal-based food products such as pasta, noodles and bread can be prepared from dough which is usually made from the basic ingredients (cereal) flour, water and optionally salt. As a result of a consumer-driven need to replace the chemical additives by more natural products, several enzymes have been developed with dough and/or cereal-based food product-improving properties, which are used in all possible combinations depending on the specific application conditions. Suitable enzymes include, for example, xylanase, starch degrading enzymes, oxidizing enzymes, fatty material splitting enzymes, protein degrading, modifying or crosslinking enzymes. Many of these enzymes are also used for treating animal feed or animal feed additives, to make them more digestible or to improve their nutritional quality. Amylases are used for the conversion of plant starches to glucose. Pectin-active enzymes are used in fruit processing, for example to increase the yield of juices, and in fruit juice clarification, as well as in other food processing steps.

There is also a growing need for improved enzymes in other industries. In the pulp and paper industry, enzymes are used to make the bleaching process more effective and to reduce the use of oxidative chemicals. In the textile industry, enzymatic treatment is often used in place

of (or in addition to) a bleaching treatment to achieve a “used” look of jeans, and can also improve the softness/feel of fabrics. When used in detergent compositions, enzymes can enhance cleaning ability or act as a softening agent. In the waste treatment industry, enzymes play an important role in changing the characteristics of the waste, for example, to become
5 more amenable to further treatment and/or for bio-conversion to value-added products.

There thus remains a need in the above-mentioned industries and others for biomass-processing enzymes, polynucleotides encoding same, and recombinant vectors and strains for expressing same.

The present description refers to a number of documents, the content of which is herein
10 incorporated by reference in their entirety.

SUMMARY OF THE INVENTION

In general, the present invention relates to a process for processing and/or degradation of biomass (e.g., cell wall deconstruction) using soluble, secreted proteins that may be isolated
15 from the fungus *Thielavia australiensis* strain ATCC 28236. Briefly, this fungal strain was cultured *in vitro* and genomic DNA along with total RNA were isolated therefrom. These nucleic acids were then used to determine/assemble fungal genomic sequences and generate a cDNA library. Bioinformatic tools were used to predict genes in the assembled genomic sequences, and those genes encoding proteins relating to biomass-degradation (e.g., cell wall
20 deconstruction) were identified based on bioinformatics (e.g., the presence of conserved domains). Sequences predicted to encode proteins which are targeted to the mitochondria or bound to the cell wall were removed. cDNA clones comprising full-length sequences predicted to encode soluble, secreted proteins relating to biomass-degradation were fully sequenced and cloned into appropriate expression vectors for protein production and characterization. The full-
25 length genomic, exonic, intronic, coding and polypeptide sequences are disclosed herein, along with corresponding putative biological functions.

The soluble, secreted, biomass degradation proteins comprise a proteome which is referred to herein as the SSBD proteome of *Thielavia australiensis*.

Accordingly, in some aspects the present invention relates to a process for degrading biomass
30 or pretreated biomass to sugars wherein an enzyme is used comprising

- a. a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 7-9;
- b. a polypeptide comprising an amino acid sequence that is at least 85%, preferably at least 90%, more preferably at least 95%, even more preferably at least 96%, 97%, 98% or

99% amino acid sequence identity to the polypeptide defined in (a);

c. a polypeptide comprising an amino acid sequence encoded by the nucleic acid sequence of any one of SEQ ID NOs: 1-6;

5 d. a polypeptide comprising an amino acid sequence encoded by a polynucleotide molecule that hybridizes under medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complement of a polynucleotide molecule comprising the nucleic acid sequence of any one of SEQ ID NOs: 1-6;

10 e. a polypeptide comprising an amino acid sequence encoded by a polynucleotide molecule having 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%, or at least 99% nucleic acid sequence identity to a polynucleotide comprising the nucleic acid sequence of any one of SEQ ID NOs: 1-6; or

15 f. a polypeptide comprising an amino acid sequence encoded by any one of the exonic nucleic acid sequences corresponding to positions: 1-92, 150-654, 815-1385, 1452-1714 of SEQ ID NO: 2; 1-246, 324-1007, 1082-1135 of SEQ ID NO: 1; 1-394, 506-789 of SEQ ID NO: 3.

In some embodiments, the above mentioned polypeptide has an activity corresponding to: cellobiohydrolase, GH6, or polysaccharide monooxygenase or GH61; or an activity according to **Table 1**.

20 In some embodiments, the above mentioned polypeptide comprises or consists of the amino acid sequence of any one of SEQ ID NOs: 7-9.

In some embodiments, the above mentioned polypeptide is a recombinant polypeptide.

In some embodiments, above mentioned polypeptide is obtainable from a fungus. In some embodiments, the fungus is from the genus *Thielavia*. In some embodiments, the fungus is *Thielavia australiensis*.

In some aspects, according to the present invention an isolated polynucleotide molecule encoding any one of the above mentioned polypeptides is used.

In some aspects, according to the present invention an isolated polynucleotide molecule is used which is:

- 30 (a) a polynucleotide molecule comprising a nucleic acid sequence encoding the polypeptide of any one of SEQ ID NOs: 7-9;
- (b) a polynucleotide molecule comprising the nucleic acid sequence of any one of 1-3;

(c) a polynucleotide molecule comprising the nucleic acid sequence of any one of SEQ ID NOs: 4-6;

(d) a polynucleotide molecule comprising any one of the exonic nucleic acid sequences corresponding to positions: 1-92, 150-654, 815-1385, 1452-1714 of SEQ ID NO: 2; 1-246, 324-1007, 1082-1135 of SEQ ID NO: 1; 1-394, 506-789 of SEQ ID NO: 3;

(e) a polynucleotide molecule comprising a nucleic acid sequence having 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%, or at least 99% nucleic acid sequence identity to any one of the polynucleotide molecules defined in (a) to (d); or

(f) a polynucleotide molecule that hybridizes under medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complement of any one of the polynucleotide molecules defined in (a) to (e).

In some embodiments, the above mentioned polynucleotide molecule is obtainable from a fungus. In some embodiments, the fungus is from the genus *Thielavia*. In some embodiments, the fungus *Thielavia australiensis*.

In some aspects, according to the present invention a vector is used comprising any one of the above mentioned polynucleotide molecules. In some embodiments, the vector comprises a regulatory sequence operatively linked to the polynucleotide molecule for expression of same in a suitable host cell. In some embodiments, the suitable host cell is a bacterial cell; a fungal cell; or a filamentous fungal cell.

In some embodiments, according to the present invention a recombinant host cell is used comprising any one of the above mentioned polynucleotide molecules or vectors. In some embodiments, according to the present invention a polypeptide is used obtainable by expressing the above mentioned polynucleotide or vector in a suitable host cell. In some embodiments, the suitable host cell is a bacterial cell; a fungal cell; or a filamentous fungal cell.

In some aspects, according to the present invention a composition is used comprising any one of the above mentioned polypeptides or the recombinant host cells. In some embodiments, the composition further comprising a suitable carrier. In some embodiments, the composition further comprises a substrate of the polypeptide. In some embodiments, the substrate is biomass.

In some aspects, according to the present invention a method is used for producing any

one of the above mentioned polypeptides, the method comprising: (a) culturing a strain comprising the above mentioned polynucleotide molecule or vector under conditions conducive for the production of the polypeptide; and (b) recovering the polypeptide. In some embodiments, the strain is a bacterial strain; a fungal strain; or a filamentous fungal strain.

5 In some aspects, according to the present invention a method is used for producing any one of the above mentioned polypeptides, the method comprising: (a) culturing the above mentioned recombinant host cell under conditions conducive for the production of the polypeptide; and (b) recovering the polypeptide.

10 In some aspects the present invention relates to the use of any one of the above mentioned polypeptides for producing ethanol. In some aspects the present invention relates to any one of the above mentioned polypeptides for producing ethanol.

In some embodiments, the above mentioned uses are in conjunction with cellulose or a cellulase.

15 In some aspects the present invention relates to the use of any one of the above mentioned polypeptides for degrading biomass or pretreated biomass. In some aspects the present invention relates to any one of the above mentioned polypeptides for degrading biomass or pretreated biomass.

20 • Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Commonly understood definitions of molecular biology terms can be found for example in Dictionary of Microbiology and Molecular Biology, 2nd ed. (Singleton et al., 1994, John Wiley & Sons, New York, NY) or The Harper Collins Dictionary of Biology (Hale & Marham, 1991, Harper Perennial, New York, NY), Rieger et al., Glossary of genetics: Classical and molecular, 5th edition, Springer-Verlag, New-York, 1991; Alberts et al., Molecular Biology of the Cell, 4th edition, Garland science, New-York, 2002; and, Lewin, Genes VII, Oxford University Press, New-York, 2000. Generally, the procedures of molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook et al., (2000, Molecular Cloning - A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratories); and Ausubel et al., (1994, Current Protocols in Molecular Biology, John Wiley & Sons, New-York).

30 • Further objects and advantages of the present invention will be clear from the description that follows.

Legends to the figures

Fig. 1 represents a schematic map of pGBFIN-49

Definitions

5 Headings, and other identifiers, e.g., (a), (b), (i), (ii), etc., are presented merely for ease of reading the specification and claims. The use of headings or other identifiers in the specification or claims does not necessarily require the steps or elements be performed in alphabetical or numerical order or the order in which they are presented.

10 In the present description, a number of terms are extensively utilized. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

15 Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one-letter nucleotide symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission.

 The use of the word “**a**” or “**an**” when used in conjunction with the term “comprising” in the claims and/or the specification may mean “one” but it is also consistent with the meaning of “one or more”, “at least one”, and “one or more than one”.

20 As used in the specification and claims, the words “**comprising**” (and any form of comprising, such as “comprise” and “comprises”), “having” (and any form of having, such as “have” and “has”), “including” (and any form of including, such as “includes” and “include”) or “containing” (and any form of containing, such as “contains” and “contain”) are inclusive or open-ended and do not exclude additional, un-recited elements or method steps.

25 The term “**about**” is used to indicate that a value includes the standard deviation of error for the device or method being employed to determine the value. In general, the terminology “about” is meant to designate a possible variation of up to 10%. Therefore, a variation of 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 % of a value is included in the term “about”.

30 The term “**DNA**” or “**RNA**” molecule or sequence (as well as sometimes the term “oligonucleotide”) refers to a molecule comprised generally of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C). In “RNA”, T is replaced by uracil (U).

 The present description refers to a number of routinely used **recombinant DNA** (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

 As used herein, “**polynucleotide**” or “**nucleic acid molecule**” refers to a polymer of

nucleotides and includes DNA (e.g., genomic DNA, cDNA), RNA molecules (e.g., mRNA), and chimeras thereof. The nucleic acid molecule can be obtained by cloning techniques or synthesized. DNA can be double-stranded or single-stranded (coding strand or non-coding strand [antisense]). Conventional deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are included in the terms "nucleic acid molecule" and "polynucleotide" as are analogs thereof (e.g., generated using nucleotide analogs, e.g., inosine or phosphorothioate nucleotides). Such nucleotide analogs can be used, for example, to prepare polynucleotides that have altered base-pairing abilities or increased resistance to nucleases. A nucleic acid backbone may comprise a variety of linkages known in the art, including one or more of sugar-phosphodiester linkages, peptide-nucleic acid bonds (referred to as "peptide nucleic acids" (PNA); Hydig-Hielsen et al., PCT Int'l Pub. No. WO 95/32305), phosphorothioate linkages, methylphosphonate linkages or combinations thereof. Sugar moieties of the nucleic acid may be ribose or deoxyribose, or similar compounds having known substitutions, e.g., 2' methoxy substitutions (containing a 2'-O-methylribofuranosyl moiety; see PCT No. WO 98/02582) and/or 2' halide substitutions. Nitrogenous bases may be conventional bases (A, G, C, T, U), known analogs thereof (e.g., inosine or others; see "The Biochemistry of the Nucleic Acids 5-36", Adams et al., ed., 11th ed., 1992), or known derivatives of purine or pyrimidine bases (see, Cook, PCT Int'l Pub. No. WO 93/13121) or "abasic" residues in which the backbone includes no nitrogenous base for one or more residues (Arnold et al., U.S. Pat. No. 5,585,481). A nucleic acid may comprise only conventional sugars, bases and linkages, as found in RNA and DNA, or may include both conventional components and substitutions (e.g., conventional bases linked via a methoxy backbone, or a nucleic acid including conventional bases and one or more base analogs).

An "**isolated nucleic acid molecule**", as is generally understood and used herein, refers to a polymer of nucleotides, and includes, but should not limited to DNA and RNA. The "isolated" nucleic acid molecule is purified from its natural *in vivo* state, obtained by cloning or chemically synthesized.

As used herein, the terms "**gene**" and "**recombinant gene**" refer to nucleic acid molecules which may be isolated from chromosomal DNA, and very often include an open reading frame encoding a protein, e.g., polypeptides of the present invention. A gene may include coding sequences, non-coding sequences, introns and regulatory sequences, as well known.

"**Amplification**" refers to any *in vitro* procedure for obtaining multiple copies ("**amplicons**") of a target nucleic acid sequence or its complement or fragments thereof. *In vitro* amplification refers to production of an amplified nucleic acid that may contain less than the complete target

region sequence or its complement. *In vitro* amplification methods include, e.g., transcription-mediated amplification, replicase-mediated amplification, polymerase chain reaction (PCR) amplification, ligase chain reaction (LCR) amplification and strand-displacement amplification (SDA including multiple strand-displacement amplification method (MSDA)). Replicase-mediated amplification uses self-replicating RNA molecules, and a replicase such as Q β -replicase (e.g., Kramer et al., U.S. Pat. No. 4,786,600). PCR amplification is well known and uses DNA polymerase, primers and thermal cycling to synthesize multiple copies of the two complementary strands of DNA or cDNA (e.g., Mullis et al., U.S. Pat. Nos. 4,683,195, 4,683,202, and 4,800,159). LCR amplification uses at least four separate oligonucleotides to amplify a target and its complementary strand by using multiple cycles of hybridization, ligation, and denaturation (e.g., EP Pat. App. Pub. No. 0320308). SDA is a method in which a primer contains a recognition site for a restriction endonuclease that permits the endonuclease to nick one strand of a hemimodified DNA duplex that includes the target sequence, followed by amplification in a series of primer extension and strand displacement steps (e.g., Walker et al., U.S. Pat. No. 5,422,252). Two other known strand-displacement amplification methods do not require endonuclease nicking (Dattagupta et al., U.S. Patent No. 6,087,133 and U.S. Patent No. 6,124,120 (MSDA)). Those skilled in the art will understand that the oligonucleotide primer sequences of the present invention may be readily used in any *in vitro* amplification method based on primer extension by a polymerase (e.g., see Kwoh et al., 1990, Am. Biotechnol. Lab. 8:14 25 and Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86, 1173 1177; Lizardi et al., 1988, BioTechnology 6:1197 1202; Malek et al., 1994, Methods Mol. Biol., 28:253 260; and Sambrook et al., 2000, Molecular Cloning - A Laboratory Manual, Third Edition, CSH Laboratories). As commonly known in the art, the oligos are designed to bind to a complementary sequence under selected conditions. The terminology "**amplification pair**" or "**primer pair**" refers herein to a pair of oligonucleotides (oligos) of the present invention, which are selected to be used together in amplifying a selected nucleic acid sequence by one of a number of types of amplification processes.

As used herein, the terms "**hybridizing**" and "**hybridizes**" are intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 60%, at least about 70%, at least about 80%, more preferably at least about 85%, even more preferably at least about 90%, more preferably at least 95%, more preferably at least 98% or more preferably at least 99% homologous to each other typically remain hybridized to each other. A preferred, non-limiting example of such hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 1X

SSC, 0.1% SDS at 50°C, preferably at 55°C, preferably at 60°C and even more preferably at 65°C. Highly stringent conditions include, for example, hybridizing at 68°C in 5x SSC/5x Denhardt's solution / 1.0% SDS and washing in 0.2x SSC/0.1% SDS at room temperature. Alternatively, washing may be performed at 42°C. The skilled artisan will know which conditions to apply for stringent and highly stringent hybridization conditions. Additional guidance regarding such conditions is readily available in the art, for example, in Sambrook et al., *supra*; and Ausubel et al., *supra* (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.). Of course, a polynucleotide which hybridizes only to a poly (A) sequence (such as the 3' terminal poly(A) tract of mRNAs), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to specifically hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The terms "**identity**" and "**percent identity**" are used interchangeably herein. For the purpose of this invention, it is defined here that in order to determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = number of identical positions/total number of positions (i.e., overlapping positions) x 100). Preferably, the two sequences are the same length. Thus, In accordance with the present invention, the term "identical" or "percent identity" in the context of two or more nucleic acid or amino acid sequences, refers to two or more sequences or subsequences that are the same, or that have a specified percentage of amino acid residues or nucleotides that are the same (e.g., 60% or 65% identity, preferably, 70-95% identity, more preferably at least 95% identity), when compared and aligned for maximum correspondence over a window of comparison, or over a designated region as measured using a sequence comparison algorithm as known in the art, or by manual alignment and visual inspection. Sequences having, for example, 60% to 95% or greater sequence identity are considered to be substantially identical. Such a definition also applies to the complement of a test sequence. Preferably, the described identity exists over a region that

is at least about 15 to 25 amino acids or nucleotides in length, more preferably, over a region that is about 50 to 100 amino acids or nucleotides in length. Those having skill in the art will know how to determine percent identity between/among sequences using, for example, algorithms such as those based on CLUSTALW computer program (Thompson Nucl. Acids Res. 2 (1994), 4673-4680) or FASTDB (Brutlag Comp. App. Biosci. 6 (1990), 237-245), as known in the art. Although the FASTDB algorithm typically does not consider internal non-matching deletions or additions in sequences, i.e., gaps, in its calculation, this can be corrected manually to avoid an overestimation of the % identity. CLUSTALW, however, does take sequence gaps into account in its identity calculations. Also available to those having skill in this art are the BLAST and BLAST 2.0 algorithms (Altschul Nucl. Acids Res. 25 (1977), 3389-3402). The BLASTN program for nucleic acid sequences uses as defaults a word length (W) of 11, an expectation (E) of 10, M=5, N=4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, and an expectation (E) of 10. The BLOSUM62 scoring matrix (Henikoff Proc. Natl. Acad. Sci., USA, 89, (1989), 10915) uses alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands. Moreover, the present invention also relates to nucleic acid molecules the sequence of which is degenerate in comparison with the sequence of an above-described hybridizing molecule. When used in accordance with the present invention the term "being degenerate as a result of the genetic code" means that due to the redundancy of the genetic code different nucleotide sequences code for the same amino acid. The present invention also relates to nucleic acid molecules which comprise one or more mutations or deletions, and to nucleic acid molecules which hybridize to one of the herein described nucleic acid molecules, which show (a) mutation(s) or (a) deletion(s). The skilled person will appreciate that all these different algorithms or programs will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when using different algorithms.

In a related manner, the terms "**homology**" or "**percent homology**", refer to a similarity between two polypeptide sequences, but take into account changes between amino acids (whether conservative or not). As well known in the art, amino acids can be classified by charge, hydrophobicity, size, etc. It is also well known in the art that amino acid changes can be conservative (e.g., they do not significantly affect, or not at all, the function of the protein). A multitude of conservative changes are known in the art, Serine for threonine, isoleucine for leucine, arginine for lysine etc., Thus the term homology introduces evolutionistic notions (e.g., pressure from evolution to a retain function of essential or important regions of a sequence, while enabling a certain drift of less important regions).

The skilled person will be aware of the fact that several different computer programs are available to determine the homology between two sequences. For instance, a comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (J. Mol. Biol. (48): 444-453 (1970)) algorithm which has been incorporated into the GAP program in the Accelrys GCG software package (available at <http://www.accelrys.com/products/gcg/>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. The skilled person will appreciate that all these different parameters will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when using different algorithms.

In yet another embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the Accelrys GCG software package (available at <http://www.accelrys.com/products/gcg/>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity two amino acid or nucleotide sequence is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989) which has been incorporated into the ALIGN program (version 2.0) (available at the ALIGN Query using sequence data of the Genestream server IGH Montpellier France <http://vega.igh.cnrs.fr/bin/align-guess.cgi>) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul et al., (1990) J. Mol. Biol. 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25(17): 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See the homepage of the National Center for Biotechnology Information at <http://www.ncbi.nlm.nih.gov/>.

By "**sufficiently complementary**" is meant a contiguous nucleic acid base sequence that is

capable of hybridizing to another sequence by hydrogen bonding between a series of complementary bases. Complementary base sequences may be complementary at each position in sequence by using standard base pairing (e.g., G:C, A:T or A:U pairing) or may contain one or more residues (including abasic residues) that are not complementary by using standard base pairing, but which allow the entire sequence to specifically hybridize with another base sequence in appropriate hybridization conditions. Contiguous bases of an oligomer are preferably at least about 80% (81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100%), more preferably at least about 90% complementary to the sequence to which the oligomer specifically hybridizes. Appropriate hybridization conditions are well known to those skilled in the art, can be predicted readily based on sequence composition and conditions, or can be determined empirically by using routine testing (see Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989) at §§ 1.90-1.91, 7.37-7.57, 9.47-9.51 and 11.47-11.57, particularly at §§ 9.50-9.51, 11.12-11.13, 11.45-11.47 and 11.55-11.57).

The present invention refers to a number of units or percentages that are often listed in sequences. For example, when referring to “at least 80%, at least 85%, at least 90%...”, or “at least about 80%, at least about 85%, at least about 90%...”, every single unit is not listed, for the sake of brevity. For example, some units (e.g., 81, 82, 83, 84, 85,... 91, 92%....) may not have been specifically recited but are considered encompassed by the present invention. The non-listing of such specific units should thus be considered as within the scope of the present invention.

Nucleic acid sequences may be detected by using hybridization with a complementary sequence (e.g., oligonucleotide probes) (see U.S. Patent Nos. 5,503,980 (Cantor), 5,202,231 (Drmanac *et al.*), 5,149,625 (Church *et al.*), 5,112,736 (Caldwell *et al.*), 5,068,176 (Vijg *et al.*), and 5,002,867 (Macevicz)). Hybridization detection methods may use an array of probes (e.g., on a DNA chip) to provide sequence information about the target nucleic acid which selectively hybridizes to an exactly complementary probe sequence in a set of four related probe sequences that differ one nucleotide (see U.S. Patent Nos. 5,837,832 and 5,861,242 (Chee *et al.*)).

A detection step may use any of a variety of known methods to detect the presence of nucleic acid by hybridization to an oligonucleotide probe. The types of detection methods in which probes can be used include Southern blots (DNA detection), dot or slot blots (DNA, RNA), and Northern blots (RNA detection). Labeled proteins could also be used to detect a particular nucleic acid sequence to which it binds (e.g., protein detection by far western technology:

Guichet et al., 1997, Nature 385(6616): 548-552; and Schwartz et al., 2001, EMBO 20(3): 510-519). Other detection methods include kits containing reagents of the present invention on a dipstick setup and the like. Of course, it might be preferable to use a detection method which is amenable to automation. A non-limiting example thereof includes a chip or other support
5 comprising one or more (e.g., an array) of different probes.

A "**label**" refers to a molecular moiety or compound that can be detected or can lead to a detectable signal. A label is joined, directly or indirectly, to a nucleic acid probe or the nucleic acid to be detected (e.g., an amplified sequence) or to a polypeptide to be detected. Direct labeling can occur through bonds or interactions that link the label to the polynucleotide or
10 polypeptide (e.g., covalent bonds or non-covalent interactions), whereas indirect labeling can occur through the use of a "**linker**" or bridging moiety, such as additional nucleotides, amino acids or other chemical groups, which are either directly or indirectly labeled. Bridging moieties may amplify a detectable signal. Labels can include any detectable moiety (e.g., a radionuclide, ligand such as biotin or avidin, enzyme or enzyme substrate, reactive group, chromophore such
15 as a dye or colored particle, luminescent compound including a bioluminescent, phosphorescent or chemiluminescent compound, and fluorescent compound).

As used herein, "**expression**" is meant the process by which a gene or otherwise nucleic acid sequence eventually produces a polypeptide. It involves transcription of the gene into mRNA, and the translation of such mRNA into polypeptide(s).

The terms "**peptide**" and "**oligopeptide**" are considered synonymous (as is commonly recognized) and each term can be used interchangeably as the context required to indicate a chain of at least two amino acids coupled by peptidyl linkages. The word "**polypeptide**" is used
20 herein for chains containing more than seven amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxyl terminus. The one-letter code of amino acids used herein is commonly known in the art and can be found in Sambrook, et al., *supra*. Sequence Listings programs can convert easily this one-letter code of amino acids sequence into a three-letter code.

The phrase "**mature polypeptide**" is defined herein as a polypeptide having biological
30 activity a polypeptide of the present invention that is in its final form, following translation and any post-translational modifications, such as N-terminal processing, C-terminal truncation, removal of signal sequences, glycosylation, phosphorylation, etc. In one embodiment, polypeptides of the present invention comprise mature of polypeptides of any one of the polypeptides disclosed herein. Mature polypeptides of the present invention can be predicted

using programs such as SignalP. The phrase "**mature polypeptide coding sequence**" is defined herein as a nucleotide sequence that encodes a mature polypeptide as defined above. As well known, some nucleotide sequences are non-coding.

As used herein, the term "**purified**" or "**isolated**" refers to a molecule (e.g., polynucleotide or polypeptide) having been separated from a component of the composition in which it was originally present. Thus, for example, an "isolated polynucleotide" or "isolated polypeptide" has been purified to a level not found in nature. A "**substantially pure**" molecule is a molecule that is lacking in most other components (e.g., 30, 40, 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99, 100% free of contaminants). By opposition, the term "**crude**" means molecules that have not been separated from the components of the original composition in which it was present. For the sake of brevity, the units (e.g., 66, 67...81, 82, 83, 84, 85, ...91, 92%....) have not been specifically recited but are considered nevertheless within the scope of the present invention.

An "**isolated polynucleotide**" or "**isolated nucleic acid molecule**" is a nucleic acid molecule (DNA or RNA) that is not immediately contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism from which it is derived. Thus, in one embodiment, an isolated nucleic acid includes some or all of the 5' non-coding (e.g., promoter) sequences that are immediately contiguous to the coding sequence. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding an additional polypeptide that is substantially free of cellular material, viral material, or culture medium (when produced by recombinant DNA techniques), or chemical precursors or other chemicals (when chemically synthesized). Moreover, an "**isolated nucleic acid fragment**" is a nucleic acid fragment that is not naturally occurring as a fragment and would not be found in the natural state.

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

The term "**variant**" refers herein to a polypeptide, which is substantially similar in structure

(e.g., amino acid sequence) to a polypeptide disclosed herein or encoded by a nucleic acid sequence disclosed herein without being identical thereto. Thus, two molecules can be considered as variants even though their primary, secondary, tertiary or quaternary structures are not identical. A variant can comprise an insertion, substitution, or deletion of one or more amino acids as compared to its corresponding native protein. A variant can comprise additional modifications (e.g., post-translational modifications such as acetylation, phosphorylation, glycosylation, sulfatation, sumoylation, prenylation, ubiquitination, etc). As used herein, the term **"functional variant"** is intended to include a variant which is sufficiently similar in both structure and function to a polypeptide disclosed herein or encoded by a nucleic acid sequence disclosed herein, to maintain at least one of its native biological activities.

As used herein, the term **"biomass"** refers to any cellulosic or lignocellulosic material and includes materials comprising cellulose, and optionally further comprising hemicellulose, lignin, starch, oligosaccharides and/or monosaccharides. Biomass may also comprise additional components, such as protein and/or lipid. Biomass may be derived from a single source, or biomass can comprise a mixture derived from more than one source; for example, biomass could comprise a mixture of corn cobs and corn stover, or a mixture of grass and leaves. Biomass includes, but is not limited to, bioenergy crops, agricultural residues, municipal solid waste, industrial solid waste, sludge from paper manufacture, yard waste, wood and forestry waste or a combination thereof. Examples of biomass include, but are not limited to, corn grain, corn cobs, crop residues such as corn husks, corn stover, grasses, wheat, wheat straw, barley, barley straw, hay, rice straw, switchgrass, waste paper, sugar cane bagasse, sorghum, soy, components obtained from milling of grains, trees, branches, roots, leaves, wood chips, sawdust, shrubs and bushes, vegetables, fruits, flowers, and animal manure or a combination thereof. Biomass that is useful for the invention may include biomass that has a relatively high carbohydrate value, is relatively dense, and/or is relatively easy to collect, transport, store and/or handle. In one embodiment of the present invention, biomass that is useful includes corn cobs, corn stover, sawdust, and sugar cane bagasse.

As used herein, the terms **"cellulosic"** or **"cellulose-containing material"** refers to a composition comprising cellulose. As used herein, the term **"lignocellulosic"** refers to a composition comprising both lignin and cellulose. Lignocellulosic material may also comprise hemicellulose. The predominant polysaccharide in the primary cell wall of biomass is cellulose, the second most abundant is hemi-cellulose, 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, cellulose is found in plant tissue primarily as an insoluble crystalline matrix of parallel glucan chains.

5 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 cellulose-containing material can be, but is not limited to, herbaceous material, agricultural residues, forestry residues, municipal solid wastes, waste paper, and pulp and paper mill residues. The cellulose-containing material can be any type of biomass including, but not limited to, wood resources, municipal solid waste, wastepaper, crops, and crop residues (e.g., see Wiselogle 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 10 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.

20 The phrase "**cellulolytic enhancing activity**" is defined herein as a biological activity which enhances the hydrolysis of a cellulose-containing material by proteins having cellulolytic activity. The term "**cellulolytic activity**" is defined herein as a biological activity which hydrolyzes a cellulose-containing material.

25 DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

POLYPEPTIDES OF THE INVENTION

In one aspect, the present invention relates to isolated polypeptides secreted by *Thielavia australiensis* (e.g., *Thielavia australiensis* strain ATCC 28236) having an activity relating to the processing or degradation of biomass (e.g., cell wall deconstruction). 30

In another aspect, the present invention relates to isolated polypeptides comprising the amino acid sequences shown in any one of SEQ ID NOs: 7-9.

In another aspect, the present invention relates to isolated polypeptides sharing a minimum threshold of amino acid sequence identity with any one of the above-mentioned polypeptides. In

specific embodiments, the present invention relates to isolated polypeptides having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% amino acid sequence identity to any one of the above-mentioned polypeptides. Other specific percentage units that have not been specifically recited here for brevity are nevertheless
5 considered within the scope of the present invention.

In another aspect, the present invention relates to a polypeptide encoded by a polynucleotide of the present invention, which includes genomic (e.g., SEQ ID NOs: 1-3), and coding (e.g., SEQ ID NOs: 4-6) nucleic acid sequences disclosed herein, polynucleotides hybridizing under medium-high, high, or very high stringency conditions with a full-length
10 complement thereof, as well as polynucleotides sharing a certain degree of nucleic acid sequence identity therewith.

In another aspect, the present invention relates to a polypeptide comprising an amino acid sequence encoded by at least one exonic nucleic acid sequence of any one of the genomic sequences corresponding to SEQ ID NOs: 1-3 (e.g., the intron or exon segments defined by the
15 exon boundaries listed in **Table 2**) or a functional part thereof.

In another aspect, the present invention relates to **functional variants** of any one of the above-mentioned polypeptides. In another embodiment, the term "functional" or "biologically active" relates to the native enzymatic (e.g., catalytic) activity of a polypeptide of the present invention. In some embodiments, the present invention relates to a polypeptide comprising a
20 biological activity of any one of the enzymes described below, or a polynucleotide encoding same.

"Carbohydrase" refers to any protein that catalyzes the hydrolysis of carbohydrates. "Glycoside hydrolase", "glycosyl hydrolase" or "glycosidase" refers to a protein that catalyzes the hydrolysis of the glycosidic bonds between carbohydrates or between a carbohydrate and a
25 non-carbohydrate residue. Endoglucanases, cellobiohydrolases, beta-glucosidases, alpha-glucosidases, xylanases, beta-xylosidases, alpha-xylosidases, galactanases, alpha-galactosidases, beta-galactosidases, alpha-amylases, glucoamylases, endo-arabinases, arabinofuranosidases, mannanases, beta-mannosidases, pectinases, acetyl xylan esterases, acetyl mannan esterases, ferulic acid esterases, coumaric acid esterases, pectin methyl esterases, and
30 chitinases are examples of glycosidases.

"Cellulase" refers to a protein that catalyzes the hydrolysis of 1,4-D-glycosidic linkages in cellulose (such as bacterial cellulose, cotton, filter paper, phosphoric acid swollen cellulose, Avicel®); cellulose derivatives (such as carboxymethylcellulose and hydroxyethylcellulose); plant lignocellulosic materials, beta-D-glucans or xyloglucans. Cellulose

is a linear beta-(1-4) glucan consisting of anhydrocellobiose units. Endoglucanases, cellobiohydrolases, and beta- glucosidases are examples of cellulases.

"**Endoglucanase**" refers to a protein that catalyzes the hydrolysis of cellulose to oligosaccharide chains at random locations by means of an endoglucanase activity.

5 "**Cellobiohydrolase**" refers to a protein that catalyzes the hydrolysis of cellulose to cellobiose via an exoglucanase activity, sequentially releasing molecules of cellobiose from the reducing or non-reducing ends of cellulose or cello- oligosaccharides. "beta-glucosidase" refers to an enzyme that catalyzes the conversion of cellobiose and oligosaccharides to glucose.

10 "**Hemicellulase**" refers to a protein that catalyzes the hydrolysis of hemicellulose, such as that found in lignocellulosic materials. Hemicelluloses are complex polymers, and their composition often varies widely from organism to organism, and from one tissue type to another. Hemicelluloses include a variety of compounds, such as xylans, arabinoxylans, xyloglucans, mamians, glucomannans, and galacto(gluco)mannans. Hemicellulose can also contain glucan, which is a general term for beta-linked glucose residues. In general, a main
15 component of hemicellulose is beta-1,4-linked xylose, a five carbon sugar. However, this xylose is often branched as beta-1,3 linkages or beta-1,2 linkages, and can be substituted with linkages to arabinose, galactose, mannose, glucuronic acid, or by esterification to acetic acid. Hemicellulolytic enzymes, i.e., hemicellulases, include both endo-acting and exo-acting
20 enzymes, such as xylanases, beta-xylosidases. alpha-xylosidases, galactanases, a- galactosidases, beta-galactosidases, endo-arabinases, arabinofuranosidases, mannanases, and beta-mannosidases. Hemicellulases also include the accessory enzymes, such as acetylesterases, ferulic acid esterases, and coumaric acid esterases. Among these, xylanases and acetyl xylan esterases cleave the xylan and acetyl side chains of xylan and the remaining xylo-oligomers are unsubstituted and can thus be hydrolysed with beta-xylosidase only. In
25 addition, several less known side activities have been found in enzyme preparations which hydrolyze hemicellulose. Accordingly, xylanases, acetylesterases and beta- xylosidases are examples of hemicellulases.

 "**Xylanase**" specifically refers to an enzyme that hydrolyzes the beta-1,4 bond in the xylan backbone, producing short xylooligosaccharides.

30 "**Beta-mannanase**" or "**endo-1,4-beta-mannosidase**" refers to a protein that hydrolyzes mannan-based hemicelluloses (mannan, glucomannan, galacto(gluco)mannan) and produces short beta-1,4-mannooligosaccharides.

 "**Mannan endo-1,6-alpha-mannosidase**" refers to a protein that hydrolyzes 1,6-alpha-mannosidic linkages in unbranched 1,6-mannans.

"**Beta-mannosidase**" (beta-1,4-mannoside mannohydrolase; EC 3.2.1.25) refers to a protein that catalyzes the removal of beta-D-mannose residues from the non-reducing ends of oligosaccharides.

5 "**Galactanase**", "**endo-beta-1,6-galactanase**" or "**arabinogalactan endo-1,4-beta-galactosidase**" refers to a protein that catalyzes the hydrolysis of endo-1,4-beta-D-galactosidic linkages in arabinogalactans.

"**Glucoamylase**" refers to a protein that catalyzes the hydrolysis of terminal 1,4-linked-D-glucose residues successively from non-reducing ends of the glycosyl chains in starch with the release of beta-D-glucose.

10 "**Beta-hexosaminidase**" or "**beta-N-acetylglucosaminidase**" refers to a protein that catalyzes the hydrolysis of terminal N-acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosamines.

15 "**Alpha-L-arabinofuranosidase**", "**alpha-N-arabinofuranosidase**", "**alpha-arabinofuranosidase**", "**arabinosidase**" or "**arabinofuranosidase**" refers to a protein that hydrolyzes arabinofuranosyl-containing hemicelluloses or pectins. Some of these enzymes remove arabinofuranoside residues from O-2 or O-3 single substituted xylose residues, as well as from O-2 and/or O-3 double substituted xylose residues. Some of these enzymes remove arabinose residues from arabinan oligomers.

20 "**Endo-arabinase**" refers to a protein that catalyzes the hydrolysis of 1,5-alpha-arabinofuranosidic linkages in 1,5-arabinans.

"**Exo-arabinase**" refers to a protein that catalyzes the hydrolysis of 1,5-alpha-linkages in 1,5-arabinans or 1,5-alpha-L arabino-oligosaccharides, releasing mainly arabinobiose, although a small amount of arabinotriose can also be liberated.

25 "**Beta-xylosidase**" refers to a protein that hydrolyzes short 1,4-beta-D-xylooligomers into xylose.

"**Cellobiose dehydrogenase**" refers to a protein that oxidizes cellobiose to cellobionolactone.

"**Chitosanase**" refers to a protein that catalyzes the endohydrolysis of beta-1,4-linkages between D-glucosamine residues in acetylated chitosan (i.e., deacetylated chitin).

30 "**Exo-polygalacturonase**" refers to a protein that catalyzes the hydrolysis of terminal alpha 1,4-linked galacturonic acid residues from non-reducing ends thus converting polygalacturonides to galacturonic acid.

"**Acetyl xylan esterase**" refers to a protein that catalyzes the removal of the acetyl groups from xylose residues. "**Acetyl mannan esterase**" refers to a protein that catalyzes the

removal of the acetyl groups from mannose residues, "**ferulic esterase**" or "**ferulic acid esterase**" refers to a protein that hydrolyzes the ester bond between the arabinose substituent group and ferulic acid. "**Coumaric acid esterase**" refers to a protein that hydrolyzes the ester bond between the arabinose substituent group and coumaric acid. Acetyl xylan esterases, ferulic acid esterases and pectin methyl esterases are examples of carbohydrate esterases.

"**Pectate lyase**" and "**pectin lyases**" refer to proteins that catalyze the cleavage of 1,4-alpha-D-galacturonan by beta-elimination acting on polymeric and/or oligosaccharide substrates (pectates and pectins, respectively).

"**Endo-1,3-beta-glucanase**" or "**laminarinase**" refers to a protein that catalyzes the cleavage of 1,3-linkages in beta-D-glucans such as laminarin or lichenin. Laminarin is a linear polysaccharide made up of beta-1,3-glucan with beta-1,6-linkages.

"**Lichenase**" refers to a protein that catalyzes the hydrolysis of lichenan, a linear, 1,3-1,4-beta-D glucan.

Rhamnogalacturonan is composed of alternating alpha-1,4-rhamnose and alpha-1,2-linked galacturonic acid, with side chains linked 1,4 to rhamnose. The side chains include Type I galactan, which is beta-1,4-linked galactose with alpha-1,3-linked arabinose substituents; Type II galactan, which is beta-1,3-1,6-linked galactoses (very branched) with arabinose substituents; and arabinan, which is alpha-1,5-linked arabinose with alpha-1,3-linked arabinose branches. The galacturonic acid substituents may be acetylated and/or methylated.

"**Exo-rhamnogalacturonanase**" refers to a protein that catalyzes the degradation of the rhamnogalacturonan backbone of pectin from the non-reducing end.

"**Rhamnogalacturonan acetylcysteine**" refers to a protein that catalyzes the removal of the acetyl groups ester-linked to the highly branched rhamnogalacturonan (hairy) regions of pectin.

"**Rhamnogalacturonan lyase**" refers to a protein that catalyzes the degradation of the rhamnogalacturonan backbone of pectin via a beta-elimination mechanism (e.g., see Pages et al., *J. Bacteriol.*, 185:4727-4733 (2003)).

"**Alpha-rhamnosidase**" refers to a protein that catalyzes the hydrolysis of terminal non-reducing alpha-L-rhamnose residues in alpha-L-rhamnosides.

Certain proteins of the present invention may be classified as "**Family 61 glycosidases**" based on homology of the polypeptides to CAZy Family GH61. Family 61 glycosidases may exhibit cellulolytic enhancing activity or endoglucanase activity. Additional information on the properties of Family 61 glycosidases may be found in U.S. Patent Application Publication Nos. 2005/0191736, 2006/0005279, 2007/0077630, and in PCT Publication No.. WO 2004/031378.

"**Esterases**" represent a category of various enzymes including **lipases**, **phospholipases**, **cutinases**, and **phytases** that catalyze the hydrolysis and synthesis of ester bonds in compounds.

The International Union of Biochemistry and Molecular Biology have developed a nomenclature for enzymes where each enzyme is described by a sequence of four numbers preceded by "EC". The first number broadly classifies the enzyme based on its mechanism. According to the naming conventions, enzymes are generally classified into six main family classes and many sub-family classes: EC 1 Oxidoreductases: catalyze oxidation/reduction reactions; EC 2 Transferases: transfer a functional group (e.g. a methyl or phosphate group); EC 3 Hydrolases: catalyze the hydrolysis of various bonds; EC 4 Lyases: cleave various bonds by means other than hydrolysis and oxidation; EC 5 Isomerases: catalyze isomerization changes within a single molecule; and EC 6 Ligases: join two molecules with covalent bonds. A number of bioinformatic tools are available to the skilled person to predict which main family class and sub-family class an enzyme molecule belongs to according to its sequence information. In some instances, certain enzymes (or family of enzymes) can be re-classified, for example, to take into account newly discovered enzyme functions or properties. Accordingly, the polypeptides/enzymes of the present invention are not meant to be limited to specific enzyme classes as they currently exist. The skilled person would know how to appropriately reclassify (and assign the appropriate functions) to the enzymes of the present invention based on the amino acid sequence information provided herein. Such reclassifications are thus within the scope of the present invention.

In some embodiments, the present invention relates to a polypeptide comprising a biological activity of any one of the enzymes (or sub-classes thereof), or a polynucleotide encoding same.

- Cellulose-hydrolyzing enzymes, including: **endoglucanases** (EC 3.2.1.4), which hydrolyze the beta-1,4-linkages between glucose units; **exoglucanases** (also known as **cellobiohydrolases** 1 and 2) (EC 3.2.1.91), which hydrolyze cellobiose, a glucose disaccharide, from the reducing and non-reducing ends of cellulose; and **beta-glucosidases** (EC 3.2.1.21), which hydrolyze the beta-1,4 glycoside bond of cellobiose to glucose;
- Proteins that enhance or accelerate the action of cellulose-degrading enzymes, including: **glycoside hydrolase family 61 (GH61) proteins** (e.g., polysaccharide monooxygenases), which enhance the action of cellulose enzymes on lignocellulose substrates;

• Enzymes that degrade or modify xylan and/or xylan-lignin complexes, including: **xylanases**, such as **endo-1,4-beta-xylanase** (EC 3.2.1.8), which catalyze the endohydrolysis of 1-4-beta-D-xylosidic linkages in xylans (or xyloglucans); **xylosidases**, such as **xylan 1,4-beta-xylosidases** (EC 3.2.1.37), which catalyze hydrolysis of 1,4-beta-D-xylans to remove successive D-xylose residues from the non-reducing terminals, and also cleaves xylobiose; **arabinosidases**, such as **alpha-arabinofuranosidases** (EC 3.2.1.55), which hydrolyze terminal non-reducing alpha-L-arabinofuranoside residues in alpha-L-arabinosides (including arabinoxylans and arabinogalactans); **alpha-glucuronidases** (EC 3.2.1.139), which hydrolyze an alpha-D-glucuronoside to the corresponding alcohol and D-glucuronate; **feruloyl esterases** (EC 3.1.1.73), which catalyzes hydrolysis of the 4-hydroxy-3-methoxycinnamoyl (feruloyl) group from an esterified sugar (which is usually arabinose in natural substrates); and **acetylxyylan esterases** (EC 3.1.1.72), which catalyze deacetylation of xylans and xylo-oligosaccharides;

• Enzymes that degrade or modify mannan, including: **mannanases**, such as **mannan endo-1,4-beta-mannosidase** (EC 3.2.1.78), which catalyze random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans and glucomannans;

• **mannosidases** (EC 3.2.1.25), which hydrolyze terminal, non-reducing beta-D-mannose residues in beta-D-mannosides; **alpha-galactosidases** (EC 3.2.1.22), which hydrolyzes terminal, non-reducing alpha-D-galactose residues in alpha-D-galactosides (including galactose oligosaccharides, galactomannans and galactohydrolase); and **mannan acetyl esterases**;

• Enzymes that degrade or modify xyloglucans, including: **xyloglucanases** such as xyloglucan-specific endo-beta-1,4-glucanase (EC 3.2.1.151), which involves endohydrolysis of 1,4-beta-D-glucosidic linkages in xyloglucan; and xyloglucan-specific exo-beta-1,4-glucanase (EC 3.2.1.155), which catalyzes exohydrolysis of 1,4-beta-D-glucosidic linkages in xyloglucan; **endoglucanases** / **cellulases**;

• Enzymes that degrade or modify glucans, including: Enzymes that degrade beta-1,4-glucan, such as **endoglucanases**; **cellobiohydrolases**; and **beta-glucosidases**;

• Enzymes that degrade beta-1,3-1,4-glucan, such as **endo-beta-1,3(4)-glucanases** (EC 3.2.1.6), which catalyzes endohydrolysis of 1,3- or 1,4-linkages in beta-D-glucans when the glucose residue whose reducing group is involved in the linkage to be hydrolyzed is itself substituted at C-3; **endoglucanases** (beta-glucanase, cellulase),

and **beta-glucosidases**;

- Enzymes that degrade or modify galactans, including: **galactanases** (EC 3.2.1.23), which hydrolyze terminal non-reducing beta-D-galactose residues in beta-D-galactosides;

- Enzymes that degrade or modify arabinans, including: **arabinanases** (EC 3.2.1.99), which catalyze endohydrolysis of 1,5-alpha-arabinofuranosidic linkages in 1,5-arabinans;

- Enzymes that degrade or modify starch, including: **amylases**, such as **alpha-amylases** (EC 3.2.1.1), which catalyze endohydrolysis of 1,4-alpha-D-glucosidic linkages in polysaccharides containing three or more 1,4-alpha-linked D-glucose units; and **glucosidases**, such as **alpha-glucosidases** (EC 3.2.1.20), which hydrolyze terminal, non-reducing 1,4-linked alpha-D-glucose residues with release of alpha-D-glucose;

- Enzymes that degrade or modify pectin, including: **pectate lyases** (EC 4.2.2.2), which carry out eliminative cleavage of pectate to give oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at their non-reducing ends; **pectin lyases** (EC 4.2.2.10), which catalyze eliminative cleavage of (1-4)-alpha-D-galacturonan methyl ester to give oligosaccharides with 4-deoxy-6-O-methyl-alpha-D-galact-4-enuronosyl groups at their non-reducing ends; **polygalacturonases** (EC 3.2.1.15), which carry out random hydrolysis of 1,4-alpha-D-galactosiduronic linkages in pectate and other galacturonans; **pectin esterases**, such as **pectin acetyl esterase** (EC 3.1.1.11), which hydrolyzes acetate from pectin acetyl esters; **alpha-arabinofuranosidases**; **beta-galactosidases**; **galactanases**; **arabinanases**; **rhamnogalacturonases** (EC 3.2.1.-), which hydrolyze alpha-D-galacturonopyranosyl-(1,2)-alpha-L-rhamnopyranosyl linkages in the backbone of the hairy regions of pectins; **rhamnogalacturonan lyases** (EC 4.2.2.-), which degrade type I rhamnogalacturonan from plant cell walls and releases disaccharide products; **rhamnogalacturonan acetyl esterases** (EC 3.1.1.-), which hydrolyze acetate from rhamnogalacturonan; and **xylogalacturonosidases** and **xylogalacturonases** (EC 3.2.1.-), which hydrolyze xylogalacturonan (xga), a galacturonan backbone heavily substituted with xylose, and which is one important component of the hairy regions of pectin;

- Enzymes that degrade or modify lignin, including: **lignin peroxidases** (EC 1.11.1.14), which oxidize lignin and lignin model compounds using hydrogen peroxide; **manganese-dependent peroxidases** (EC 1.11.1.13), which oxidizes lignin and lignin

model compounds using Mn^{2+} and hydrogen peroxide; **versatile peroxidases** (EC 1.11.1.16), which oxidize lignin and lignin model compounds using an electron donor and hydrogen peroxide and combines the substrate-specificity characteristics of the two other ligninolytic peroxidases: manganese peroxidase (EC 1.11.1.13) and lignin peroxidase (EC 1.11.1.14); and **laccases** (EC 1.10.3.2), a group of multi-copper proteins of low specificity acting on both o- and p-quinols, and often acting also on lignin; and

- Enzymes acting on chitin, including: **chitinases** (EC 3.2.1.14), which catalyze random hydrolysis of N-acetyl-beta-D-glucosaminide 1,4-beta-linkages in chitin and chitodextrins; and **hexosaminidases**, such as **beta-N-acetylhexosaminidase** (EC 3.2.1.52), which hydrolyzes terminal non-reducing N-acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides.

In another embodiment, the present invention includes the polypeptides and their corresponding activities as defined in **Table 1**, as well as functional variants thereof.

As alluded to above, the term “**functional variant**” as used herein is intended to include a polypeptide which is sufficiently similar in structure and function to any one of the above-mentioned polypeptides (without being identical thereto) to maintain at least one of its native biological activities. In another embodiment, a functional variant can comprise an insertion, substitution, or deletion of one or more amino acids as compared to its corresponding native protein. In another embodiment, a functional variant can comprise additional modifications (e.g., post-translational modifications such as acetylation, phosphorylation, glycosylation, sulfatation, sumoylation, prenylation, ubiquitination, etc).

In another embodiment, functional variants of the present invention can contain one or more **conservative substitutions** of a polypeptide sequence disclosed herein. Such modifications can be carried out routinely using site-specific mutagenesis. The term “conservative substitution” is intended to indicate a substitution in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acids having similar side chains are known in the art and include amino acids with basic side chains (e.g., lysine, arginine and histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagines, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

In another embodiment, functional variants of the present invention can contain one or more insertions, deletions or truncations of non-essential amino acids. As used herein, a “**non-**

essential amino acid" is a residue that can be altered in a polypeptide of the present invention without substantially altering its biological function. For example, amino acid residues that are conserved among the proteins of the present invention having similar biological activities (and their orthologs) are predicted to be particularly unamenable to alteration.

5 In another embodiment, functional variants can include **functional fragments** (i.e., biologically active fragments) of any one of the polypeptide sequences disclosed herein. Such fragments include fewer amino acids than the full length protein from which they are derived, but exhibit at least one biological activity of the corresponding full-length protein. Typically, biologically active fragments comprise a domain or motif with at least one activity of the full-
10 length protein. A biologically active fragment of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the biological activities of the native form of a polypeptide of the present invention.

15 In another embodiment, the present invention includes other functional variants of the polypeptides disclosed herein, which can be identified by techniques known in the art. For example, functional variants can be identified by screening combinatorial libraries of mutants (e.g., truncation mutants), of polypeptides of the present invention for biological activity. In another embodiment, a variegated library of variants can be generated by combinatorial
20 mutagenesis at the nucleic acid level. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods that can be used to produce libraries of potential variants of the
25 polypeptides of the present invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (e.g., see Narang (1983) Tetrahedron 39:3; Itakura et al., (1984) Annu. Rev. Biochem. 53:323; Itakura et al., (1984) Science 198:1056; Ike et al., (1983) Nucleic Acid Res. 11:477).

30 In addition, libraries of fragments of the coding sequence of a polypeptide of the present invention can be used to generate a variegated population of polypeptides for screening a subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include

sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

5 Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations of truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library
10 of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of polypeptides of the present invention (Arkin and Yourvan (1992) Proc. Natl.
15 Acad. Sci. USA 89:7811-7815; Delgrave et al., (1993) Protein Engineering 6(3): 327-331).

 In another embodiment, functional variants of the present invention can encompass orthologs of the genes and polypeptides disclosed herein. Orthologs of the polypeptides disclosed herein include proteins that can be isolated from other strains or species and possess a similar or identical biological activity. Such orthologs can be identified as comprising an amino
20 acid sequence that is substantially homologous (shares a certain degree of amino acid sequence identity) with the polypeptides disclosed herein. As used herein, the expression “**substantially homologous**” refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with similar side chain) amino acids or nucleotides to a second amino acid or nucleotide sequence such that the first and the
25 second amino acid or nucleotide sequences have a common domain. For example, amino acid or nucleotide sequences which contain a common domain having at least 85%, 86%, 87%, 88%, 89%, 90%, 91% 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity are defined herein as sufficiently identical.

 In another embodiment, the present invention includes improved proteins derived from the
30 polypeptides of the present invention. Improved proteins are proteins wherein at least one biological activity is improved. Such proteins may be obtained by randomly introducing mutations along all or part of the coding sequences of the polypeptides of the present invention such as by saturation mutagenesis, and the resulting mutants can be expressed recombinantly and screened for biological activity. For instance, the art provides for standard assays for

measuring the enzymatic activity of the resulting protein and thus improved proteins may be selected.

Recovery and purification

5 In another aspect, polypeptides of the present invention may be present alone (e.g., in an isolated or purified form), within a composition (e.g., an enzymatic composition for carrying out an industrial process), or in an appropriate host. In one embodiment, polypeptides of the present invention can be recovered and purified from cell cultures (e.g., recombinant cell cultures) by methods known in the art. In another embodiment, high performance liquid
10 chromatography ("HPLC") can be employed for the purification.

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

Fusion proteins

20 In another aspect, the present invention includes **fusion proteins** comprising a polypeptide of the present invention or a functional variant thereof, which is operatively linked to one or more unrelated polypeptide (e.g., heterologous amino acid sequences). "**Unrelated polypeptides**" or "**heterologous polypeptides**" or "**heterologous sequences**" refer to polypeptides or sequences which are usually not present close to or fused to one of the polypeptides of the present invention. Such "unrelated polypeptides" or "heterologous polypeptides" having amino
25 acid sequences corresponding to proteins which are not substantially homologous to the polypeptide sequences disclosed herein. Such "unrelated polypeptides" can be derived from the same or a different organism. In one embodiment, a fusion protein of the present invention comprises at least two biologically active portions or domains of polypeptide sequences
30 disclosed herein. In the context of fusion proteins, the term "operatively linked" is intended to indicate that all of the different polypeptides are fused in-frame to each other. In another embodiment, an unrelated polypeptide can be fused to the N terminus or C terminus of a polypeptide of the present invention.

In another embodiment, a polypeptide of the present invention can be fused to a protein

which enables or facilitates recombinant protein purification and/or detection. For example, a polypeptide of the present invention can be fused to a protein such as glutathione S-transferase (GST), and the resulting fusion protein can then be purified/detected through the high affinity of GST for glutathione.

5 Fusion proteins of the present invention can be produced by standard recombinant DNA techniques. For example, DNA fragments encoding different polypeptide sequences can be ligated together in frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to
10 avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers, which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (e.g.,
15 see *Current Protocols in Molecular Biology*, eds. Ausubel et al., John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the present invention can be cloned into such an expression vector so that the fusion moiety is linked in-frame to the polypeptide of interest.

Signal sequences

20 In another embodiment, a polypeptide of the present invention can be fused to a heterologous signal sequence (e.g., at its N terminus) to facilitate its isolation, expression and/or secretion from certain host cells (e.g., mammalian and yeast host cells). Signal sequences are
25 typically characterized by a core of hydrophobic amino acids, which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides may contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway.

30 For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the *phoA* secretory signal (Sambrook et al., *supra*) and the protein A

secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

The signal sequence can direct secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by known methods. In another embodiment, a signal sequence can be linked to a fusion protein of the present invention to facilitate detection, purification, and/or recovery thereof. For example, the sequence encoding a fusion protein of the present invention may be fused to a marker sequence, such as a sequence encoding a peptide, which facilitates purification of the fused polypeptide. In another embodiment, the marker sequence can be a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. In another embodiment, the HA tag is another peptide useful for purification, which corresponds to an epitope derived of influenza hemagglutinin protein, which has been described by Wilson et al., Cell 37:767 (1984), for instance.

POLYNUCLEOTIDES

The nucleic acid sequences of the genes disclosed herein were determined by sequencing cDNA clones, mRNA transcripts, or genomic DNA obtained from *Thielavia australiensis* ATCC 28236. In one aspect, a polynucleotide of the present invention may be obtained from a fungi, in particular from *Thielavia*, and preferably from *Thielavia australiensis*.

In another aspect, the present invention relates to polynucleotides encoding a polypeptide of the present invention, including functional variants thereof. In one embodiment, polynucleotides of the present invention comprise the **coding nucleic acid sequence** of any one of SEQ ID NOs: 4-6, or as set forth in **Table 1**.

In another aspect, the present invention relates to genomic DNA sequences corresponding to the above mentioned coding sequences. In one embodiment, polynucleotides of the present invention comprise the **genomic nucleic acid sequence** of any one of SEQ ID NOs: 1-3; or as set forth in **Table 1**.

In another aspect, the present invention relates to a polynucleotide comprising at least one intronic or exonic nucleic acid sequence of any one of the genomic sequences corresponding to SEQ ID NOs: 1-3; (e.g., the intron or exon segments defined by the exon boundaries listed in **Table 2**). Although only the positions of the exons are defined in **Table 2**, a person of skill in the art would readily be able to determine the positions of the corresponding introns in view of this

information. In some embodiments, polynucleotides comprising at least one these intronic segments are within the scope of the present invention.

In yet another aspect, the present invention relates to a polynucleotide comprising at least one exonic nucleic acid sequence comprised within SEQ ID NOs: 1-3 or as set forth in **Table 2**.

5 In another aspect, the present invention relates to isolated polynucleotides sharing a minimum threshold of nucleic acid sequence identity with any one of the above-mentioned polynucleotides. In specific embodiments, the present invention relates to isolated polynucleotides having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% nucleic acid sequence identity to any one of the above-mentioned
10 polynucleotides. Other specific percentage units that have not been specifically recited here for brevity are nevertheless considered within the scope of the present invention. Polynucleotides having the aforementioned thresholds of nucleic acid sequence identity can be created by introducing one or more nucleotide substitutions, additions or deletions into the coding nucleotide sequences of the present invention such that one or more amino acid substitutions,
15 deletions or insertions are introduced into the encoded polypeptide. Such mutations may be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis.

In another aspect, the present invention relates to a polynucleotide that hybridizes (or is hybridizable) under medium-high stringency conditions, high stringency conditions, or very high
20 stringency conditions with the full-length complement of any one of the polynucleotides defined above.

As used herein, "**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
25 standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 2X SSC, 0.2% SDS at 45°C.

As used herein, "**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
30 standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 2X SSC, 0.2% SDS at 50°C.

As used herein, "**medium stringency conditions**" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SOS, 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 2X SSC, 0.2% SDS at 55°C.

As used herein, "**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
5 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 2X SSC, 0.2% SDS at 60°C.

As used herein, "**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
10 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 2X SSC, 0.2% SDS at 65°C.

As used herein, "**very 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
15 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 2X SSC, 0.2% SDS at 70°C.

In one embodiment, a polynucleotide of the present invention (or a fragment thereof) can be isolated using the sequence information provided herein in conjunction with standard molecular
20 biology techniques (e.g., as described in Sambrook et al., *supra*. For example, suitable hybridization oligonucleotides (e.g., probes or primers) can be designed using all or a portion of the nucleic acid sequences disclosed herein and prepared by standard synthetic techniques (e.g., using an automated DNA synthesizer). The oligonucleotides can be employed in hybridization and/or amplification reactions, for example, to amplify a template of cDNA, mRNA
25 or genomic DNA, according to standard PCR techniques. A polynucleotide so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

In another aspect, the present invention relates to polynucleotides encoding functional variants of any one of the polypeptides of the present invention, including a biologically active fragment or domain thereof.

30 In another aspect, the present invention can include nucleic acid molecules (e.g., oligonucleotides) sufficient for use as primers and/or hybridization probes to amplify, sequence and/or identify nucleic acid molecules encoding a polypeptide of the present invention or fragments thereof. In some embodiments, the present invention relates to polynucleotides (e.g., oligonucleotides) that comprise, span, or hybridize specifically to exon-exon or exon-intron

junctions of the genomic sequences identified herein, such as those defined in **Table 2**. Designing such polynucleotides/oligonucleotides would be within the grasp of a person of skill in the art in view of the target sequence information disclosed herein and are thus encompassed by the present invention.

5 In another aspect, the present invention relates to polynucleotides comprising silent mutations or mutations that do not significantly alter the biological function of the encoded polypeptide. Guidance concerning how to make phenotypically silent amino acid substitutions is provided for example in Bowie et al., Science 247:1306-1310 (1990) and in the references cited therein. Furthermore, it will be apparent for the skilled person that DNA sequence
10 polymorphisms of the genes disclosed herein may exist within a given population, which may differ from the sequences disclosed herein. Such genetic polymorphisms may exist in cells from different populations or within a population due to natural allelic variation. Accordingly, in one embodiment, the present invention can include natural allelic variants and homologs of polynucleotides disclosed herein.

15 In another aspect, polynucleotides of the present invention can comprise only a portion or a fragment of the nucleic acid sequences disclosed herein. Although such polynucleotides may not encode a functional polypeptide of the present invention, they are useful for example as probes or primers in hybridization or amplification reactions. Exemplary uses of such polynucleotides include: (1) isolating a gene (as allelic variant thereof) from cDNA library; (2) *in*
20 *situ* hybridization (e.g., FISH) to metaphase chromosomal spreads to provide precise chromosomal location of the gene as described in Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York (1988); (3) Northern blot analysis for detecting expression of mRNA corresponding to a polypeptide disclosed herein, or a homolog, ortholog or variant thereof, in specific tissues and/or cells; and (4) probes and primers that can
25 be used as a diagnostic tool to analyze the presence of a nucleic acid hybridizable to a polynucleotide disclosed herein in a given biological (e.g., tissue) sample. It would be within the grasp of a skilled person to design specific oligonucleotides in view of the nucleic acid sequences disclosed herein. Oligonucleotides typically comprise a region of nucleotide sequence that hybridizes (preferably under highly stringent conditions) to at least 10, 11, 12, 13,
30 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 37, 39, 40, 50, 60, 70, 80, 90 or 100 contiguous nucleotides of a polynucleotide of the present invention. In one embodiment, such oligonucleotides can be used for identifying and/or cloning other family members, as well as orthologs from other species. In another embodiment, the oligonucleotide can be attached to a detectable label (e.g., a radioisotope, a fluorescent

compound, an enzyme, or an enzyme cofactor). Such oligonucleotides can also be used as part of a diagnostic method or kit for identifying cells which express a polypeptide of the present invention.

As would be understood by the skilled person, full-length complements of any one of the polynucleotides of the present invention are also encompassed. In one embodiment, the full-length complements are antisense molecules with respect to the coding strands of polynucleotides of the present invention, which hybridize (preferably under highly stringent conditions) to at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 37, 39, 40, 50, 60, 70, 80, 90 or 100 contiguous nucleotides to a polynucleotide of the present invention.

Sequencing errors

The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The specific sequences disclosed herein can be readily used to isolate the corresponding complete genes from the organism sequenced herein, which in turn can easily be subjected to further sequence analyses thereby identifying sequencing errors.

Unless otherwise indicated, all nucleotide sequences disclosed herein were determined by sequencing using an automated DNA sequencer, and all amino acid sequences of polypeptides disclosed herein were predicted by translation based on the genetic code. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The person skilled in the art is capable of identifying such erroneously identified bases and knows how to correct such errors.

VECTORS

Another aspect of the invention pertains to vectors (e.g., expression vectors), containing a polynucleotide encoding a polypeptide of the present invention.

As used herein, the term “**vector**” includes a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “**expression vectors**”. In general, expression vectors useful in recombinant DNA techniques are often in the form of plasmids. The terms “**plasmid**” and “**vector**” can be used interchangeably herein as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

In one embodiment, recombinant expression vectors of the invention can comprise a polynucleotide of the present invention in a form suitable for expression of the polynucleotide in a host cell, which means that the recombinant expression vector includes one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, “**operatively linked**” is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term “**regulatory sequence**” is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signal). Such regulatory sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in a certain host cell (e.g., tissue-specific regulatory sequences). It will be appreciated by those

skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the present invention can be introduced into host cells to thereby produce proteins or peptides, encoded by polynucleotides as described herein (e.g., polypeptides of the present invention).

In another embodiment, recombinant expression vectors of the present invention can be designed for expression of polypeptides of the present invention in prokaryotic or eukaryotic cells. For example, these polypeptides can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel *supra*). In another embodiment, recombinant expression vectors of the present invention can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

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

For expression, a DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled person. In a specific embodiment, promoters are preferred that are capable of directing a high expression level of biologically active polypeptides of the present invention (e.g., lignocellulose active proteins) from fungi. Such promoters are known in the art. The expression constructs may contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, transduction, infection, lipofection,

cationic lipid-mediated transfection or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al., (Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), Davis et al., Basic Methods in Molecular Biology (1986) and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methatrexate. A polynucleotide encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a polypeptide of the present invention, or on a separate vector. Cells stably transfected with a polynucleotide of the present invention can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

Expression of proteins in prokaryotes is often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, e.g., to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (1) to increase expression of recombinant protein; (2) to increase the solubility of the recombinant protein; and (3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein.

Vectors preferred for use in bacteria are for example disclosed in WO-A1-2004/074468. Other suitable vectors will be readily apparent to the skilled artisan. Known bacterial promoters suitable for use in the present invention include the promoters disclosed in WO-A1-2004/074468.

As indicated, the expression vectors will preferably contain selectable markers. Such markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture and antibiotic resistance (e.g., tetracycline or ampicillin) for culturing in *E. coli* and other bacteria. Representative examples of appropriate host include bacterial cells, such as *E. coli*, *Streptomyces*, *Salmonella typhimurium* and certain *Bacillus* species; fungal cells such as

Aspergillus species, for example *A. niger*, *A. oryzae* and *A. nidulans*, yeast cells such as *Kluyveromyces*, for example *K. lactis* and/or *Pichia*, for example *P. pastoris*; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9; animal cells such as CHO, COS and Bowes melanoma; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signal may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals. In an embodiment, a polypeptide of the present invention may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification and/or detection.

HOST CELLS

In another aspect, the present invention features cells, e.g., transformed host cells or recombinant host cells that contain a polynucleotide or vector of the present invention. A "**transformed cell**" or "**recombinant cell**" is a cell into which (or into an ancestor of which) has been introduced a polynucleotide or vector of the invention by means of recombinant DNA techniques. Both prokaryotic and eukaryotic cells are included, e.g., bacteria, fungi, yeast, and the like, especially preferred are cells from filamentous fungi, in particular the strain from which the polynucleotide and polypeptide sequences disclosed herein were derived.

In one embodiment, a cell of the present invention is typically not a wild-type strain or a naturally-occurring cell. Host cells of the present invention can include, but are not limited to: fungi (e.g., *Aspergillus niger*, *Trichoderma reesii*, *Myceliophthora thermophila* and *Talaromyces*

emersonii); yeasts (e.g., *Saccharomyces cerevisiae*, *Yarrowia lipolytica* and *Pichia pastoris*); bacteria (e.g., *Escherichia coli* and *Bacillus sp.*); and plants (e.g., *Nicotiana benthamiana*, *Nicotiana tabacum* and *Medicago sativa*).

5 In another embodiment, a polynucleotide (or a polynucleotide which is comprised within a vector) may be homologous or heterologous with respect to the cell into which it is introduced. In this context, a polynucleotide is homologous to a cell if the polynucleotide naturally occurs in that cell. A polynucleotide is heterologous to a cell if the polynucleotide does not naturally occur in that cell. Accordingly, in an embodiment, the present invention relates to a cell which comprises a heterologous or a homologous sequence corresponding to any one of
10 the polynucleotides or polypeptides disclosed herein.

In another embodiment, a host cell can be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in a specific, desired fashion. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may facilitate optimal functioning of the protein. Various host cells have characteristic
15 and specific mechanisms for post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems familiar to those of skill in the art can be chosen to ensure the desired and correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product can
20 be used. Such host cells are well known in the art.

In another embodiment, host cells can also include, but are not limited to, mammalian cell lines such as CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and choroid plexus cell lines. If desired, a stably transfected cell line can produce the polypeptides of the present invention. A number of vectors suitable for stable transfection of mammalian cells
25 are available to the public, methods for constructing such cell lines are also publicly known, e.g., in Ausubel et al., (*supra*).

In another embodiment, the present invention relates to methods of inhibiting the expression of a polypeptide of the present invention in a host cell, comprising administering to the cell or expressing in the cell a double-stranded RNA (dsRNA) molecule (or a molecule
30 comprising region of double-strandedness), wherein the dsRNA comprises a subsequence of a polynucleotide of the present invention. In a preferred aspect, the dsRNA is about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more duplex nucleotides in length. The dsRNA is preferably a small interfering RNA (siRNA) or a micro RNA (miRNA). In a preferred aspect, the dsRNA is small interfering RNA (siRNAs) for inhibiting transcription. In another preferred aspect, the

dsRNA is micro RNA (miRNAs) for inhibiting translation. The present invention also relates to such double-stranded RNA (dsRNA) molecules, comprising a portion of the mature polypeptide coding sequence of any one of the coding sequences of the polypeptides disclosed herein of inhibiting expression of that polypeptide in a cell. While the present invention is not limited by any particular mechanism of action, the dsRNA can enter a cell and cause the degradation of a single-stranded RNA (ssRNA) of similar or identical sequences, including endogenous mRNAs. When a cell is exposed to dsRNA, mRNA from the homologous gene is selectively degraded by a process called RNA interference (RNAi). The dsRNAs of the present invention can be used in gene-silencing methods. In one aspect, the invention relates to methods to selectively degrade RNA using the dsRNAi's of the present invention. The process may be practiced *in vitro*, *ex vivo* or *in vivo*. In one aspect, the dsRNA molecules can be used to generate a loss-of-function mutation in a cell, an organ or an organism. Methods for making and using dsRNA molecules to selectively degrade RNA are well known in the art, see, for example, U.S. Patent No. 6,506,559; U.S. Patent No. 6,511,824; U.S. Patent No. 6,515,109; and U.S. Patent No. 6,489,127. In some instances, new phylogenetic analyses of fungal species have resulted in taxonomic reclassifications. For example, following their phylogenetic studies reported in van den Brink et al., ("Phylogeny of the industrial relevant, thermophilic genera *Myceliophthora* and *Corynascus*", Fungal Diversity (2012), 52:197-207), the authors proposed renaming all existing *Corynascus* species to *Myceliophthora*. Such changes in taxonomic classification are within the scope of the present invention and, regardless of future reclassifications, a person of skill in the art would be able to identify the organism used to determine the sequences disclosed herein for example based on the strain's accession number (ATCC 28236).

It should be understood herein that the level of expression of polypeptides of the present invention could be modified by adapting the codon usage ratio of a sequence of the present invention to that of the host or hosts in which it is meant to be expressed. This adaptation and the concept of codon usage ratio are all well known in the art.

Antibodies

In another aspect, the present invention relates to an isolated binding agent capable of selectively binding to a polypeptide of the present invention. Suitable binding agents may be selected from an antibody, an antigen binding fragment, or a binding partner. In one embodiment, the binding agent selectively binds to an amino acid sequence selected from **Table 1**, including to any fragment of any of the above sequences comprising at least one antibody binding epitope.

According to the present invention, the phrase "**selectively binds to**" refers to the ability of an antibody, antigen binding fragment or binding partner of the present invention to preferentially bind to specified proteins. More specifically, the phrase "**selectively binds**" refers to the specific binding of one protein to another (e.g., an antibody, fragment thereof, or binding partner to an antigen), wherein the level of binding, as measured by any standard assay (e.g., an immunoassay), is statistically significantly higher than the background control for the assay. For example, when performing an immunoassay, controls typically include a reaction well/tube that contain antibody or antigen binding fragment alone (i.e., in the absence of antigen), wherein an amount of reactivity (e.g., non-specific binding to the well) by the antibody or antigen binding fragment thereof in the absence of the antigen is considered to be background. Binding can be measured using a variety of methods standard in the art including enzyme immunoassays (e.g., ELISA), immunoblot assays, etc.).

Antibodies are characterized in that they comprise immunoglobulin domains and as such, they are members of the immunoglobulin superfamily of proteins. An antibody of the invention includes polyclonal and monoclonal antibodies, divalent and monovalent antibodies, bi- or multi-specific antibodies, serum containing such antibodies, antibodies that have been purified to varying degrees, and any functional equivalents of whole antibodies. Isolated antibodies of the present invention can include serum containing such antibodies, or antibodies that have been purified to varying degrees. Whole antibodies of the present invention can be polyclonal or monoclonal. Alternatively, functional equivalents of whole antibodies, such as antigen binding fragments in which one or more antibody domains are truncated or absent (e.g., Fv, Fab, Fab', or F(ab)₂ fragments), as well as genetically-engineered antibodies or antigen binding fragments thereof, including single chain antibodies or antibodies that can bind to more than one epitope (e.g., bi-specific antibodies), or antibodies that can bind to one or more different antigens (e.g., bi- or multi-specific antibodies), may also be employed in the invention. Methods for the generation and production of antibodies are well known in the art.

Monoclonal antibodies may be produced according to the methodology of Kohler and Milstein (Nature 256:495-497, 1975). Non-antibody polypeptides, sometimes referred to as binding partners, may be designed to bind specifically to a protein of the invention. Examples of the design of such polypeptides, which possess a prescribed ligand specificity are given in Beste et al., (Proc. Nat'l Acad. Sci. 96:1898-1903, 1999). In one embodiment, a binding agent of the invention is immobilized on a substrate such as: artificial membranes, organic supports, biopolymer supports and inorganic supports such as for use in a screening assay.

In some embodiment, antibodies and binding agents specifically binding to polypeptides of

the present invention may be produced and used even in absence of knowledge of the precise biological function of the polypeptide. Such antibodies and binding agent may be useful, for example, as diagnostic, classification, and/or research tools.

5 **COMPOSITIONS AND USES**

In another aspect, the present invention relates to composition comprising one or more polypeptides or polynucleotides of the present invention. In one embodiment, the compositions are enriched in such a polypeptide. The term "**enriched**" indicates that the biological activity (e.g., biomass degradation or processing) of the composition has been increased, e.g., with an enrichment factor of at least 1.1. The composition may comprise a polypeptide of the present invention as the major component, e.g., a mono-component composition. Alternatively, the composition may comprise multiple enzymatic activities (e.g., those described herein).

The polypeptide compositions may be prepared in accordance with methods known in the art and may be in the form of a liquid or a dry composition. For instance, the polypeptide composition may be in the form of a granulate or a microgranulate. The polypeptide to be included in the composition may be stabilized in accordance with methods known in the art. Examples are given below of preferred uses of the polypeptide compositions of the present invention. The dosage of the polypeptide composition of the invention and other conditions under which the composition is used may be determined on the basis of methods known in the art.

In another aspect, the present invention relates to the use of the polypeptides (e.g., enzymes) of the present invention a number of industrial and other processes. Despite the long term experience obtained with these processes, there remains a need for improved polypeptides and enzymes featuring one or more significant advantages over those presently used. Depending on the specific application, these advantages can include aspects such as lower production costs, higher specificity towards the substrate, greater synergies with existing enzymes, less antigenic effect, less undesirable side activities, higher yields when produced in a suitable microorganism, more suitable pH and temperature ranges, better properties of the final product, and food grade or kosher aspects. In various embodiments, the present invention seeks to provide one or more of these advantages, or others.

Biomass processing or degradation

In another aspect, the polypeptides of the present invention may be used in new or improved methods for enzymatically degrading or converting plant cell wall polysaccharides

from biomass into various useful products. In addition to cellulose and hemicellulose, plant cell walls contain associated pectins and lignins, the removal of which by enzymes of the current invention can improve accessibility to cellulases and hemicellulases, or which can themselves be converted to useful products. Therefore the polypeptides of the present invention may be used to degrade biomass or pretreated biomass to sugars. These sugars may be used as such or may be, for example, fermented into ethanol.

Usually, biomass must be subjected to pre-treatment in order to make the cellulose more accessible. Accordingly, in one embodiment, polypeptides of the present invention may be used in improved methods for the processing of pretreated biomass. Pretreatment technologies may involve chemical, physical, or biological treatments. Examples of pre-treatment technologies include but are not limited to: steam explosion; ammonia; acid hydrolysis; alkaline hydrolysis; solvent extraction; crushing; milling; etc.

One example of a product produced from biomass is bioethanol. Bioethanol is usually produced by the fermentation of glucose to ethanol by yeasts such as *Saccharomyces cerevisiae*: in addition to ethanol, other chemicals may be synthesized starting from glucose. Ethanol, today, is produced mostly from sugars or starches, obtained from sugar cane, fruits and grains. In contrast, cellulosic ethanol is obtained from cellulose, the main component of wood, straw and much of the plants. Sources of biomass for cellulosic ethanol production comprise agricultural residues (e.g., leftover crop materials from stalks, leaves, and husks of corn plants), forestry wastes (e.g., chips and sawdust from lumber mills, dead trees, and tree branches), energy crops (e.g., dedicated fast-growing trees and grasses such as switch grass), municipal solid waste (e.g., household garbage and paper products), food processing and other industrial wastes (e.g., black liquor, paper manufacturing by-products, etc.).

Plant biomass is a mixture of plant polysaccharides, including cellulose, hemicelluloses, and pectin, together with the structural polymer, lignin. Glucose is released from cellulose by the action of mixtures of enzymes, including: endoglucanases, exoglucanases (cellobiohydrolases 1 and 2) and beta-glucosidases. Efficient large-scale conversion of cellulosic materials by such mixtures may require the full complement of enzymes, and can be enhanced by the addition of enzymes that attack the other plant cell wall components (e.g., hemicelluloses, pectins, and lignins), as well as chemical linkages between these components. Hence, polypeptides of the present invention that are highly expressed, or have high specific activity, stability, or resistance to inhibitors may improve the efficiency of the process, and lower enzyme costs. It would be an advantage to the art to improve the degradation and conversion of plant cell wall polysaccharides by composing cellulase mixtures using cellulase enzymes with

such properties. Furthermore, polypeptides of the present invention that are able to function at extremes of pH and temperature are desirable, both since improved enzyme robustness decreases costs, and because enzymes that function at high temperature will allow high processing temperatures under high substrate consistency conditions that decrease viscosity and thus improve yields.

Glycoside hydrolases from the family GH61 are known to stimulate the activity of cellulose cocktails on lignocellulosic substrates and are thus considered to exhibit cellulose-enhancing activity (Harris et al., Biochemistry 49, 3305 (2010)). They have no known enzymatic activities of their own. Enhancement of cellulase cocktail efficiency by GH61 proteins of the present invention may contribute to lowering the costs of cellulase enzymes used for the production of glucose from plant cell biomass, as described above. GH61 (glycoside hydrolase family 61 or sometimes referred to as EGIV) proteins are oxygen-dependent polysaccharide monooxygenases (PMO's) according to the latest literature. Often in the literature, these proteins are mentioned as enhancing the action of cellulases on lignocellulose substrates. GH61 was originally classified as an endoglucanase, based on the measurement of very weak endo-1,4- β -d-glucanase activity in one family member. The term "**GH61**" as used herein, is to be understood as a family of enzymes, which share common conserved sequence portions and foldings to be classified in family 61 of the well-established CAZY GH classification system (<http://www.cazy.org/GH61.html>). The glycoside hydrolase family 61 is a member of the family of glycoside hydrolases EC 3.2.1. GH61 is used herein as being part of the cellulases.

Enzymatic hydrolysis of plant hemicellulose yields 5-carbon sugars that either may be fermented to ethanol by some species of yeast, or converted to other types of chemical products. Enzymatic deconstruction of hemicellulose is also known to improve the accessibility of plant cell wall cellulose to cellulase enzymes for the production of glucose from lignocellulosic materials. Hemicellulase enzymes of the present invention that enhance glucose production from lignocellulose would find utility in the bioethanol industry and in other process that rely on glucose or pentose streams from lignocellulose.

Lignin is composed of methoxylated phenyl-propane units linked by ether linkages and carbon-carbon bonds. The chemical composition of lignin may, depending on species, include guaiacyl, 4-hydroxyphenyl, and syringyl groups. Enzymatic modification of lignin by the polypeptides of the present invention can be used for the production of structural materials from plant biomass, or alternatively improve the accessibility of plant cellulose and hemicelluloses to cellulase enzymes for the release of glucose from biomass as described above. Enzymes that degrade the lignin component of lignocellulose include lignin peroxidases, manganese-

dependent peroxidases, versatile peroxidases, and laccases (Vicuna et al., 2000, Molecular Biotechnology 14: 173-176; Broda et al., 1996, Molecular Microbiology 19: 923-932). In some embodiments, polypeptides of the present invention may also, in certain instances, be active in the decolorization of industrial dyes, and thus useful for the treatment and detoxification of chemical wastes.

In another embodiment, pectin-degrading polypeptides of the present invention can also enhance the action of cellulases on plant biomass by improving the accessibility of cellulase to the cellulose component of lignocellulose.

In another embodiment, polypeptides of the present invention may also be useful in other applications for hydrolyzing non-starch polysaccharide (NSP).

In another embodiment, esterases of the present invention can be useful in the bioenergy industry such as for the production of biodiesel and hydrolysis of hemicellulose.

In another embodiment, the present invention relates to methods for degrading or converting a cellulose-containing material, comprising: treating the cellulose-containing material with an effective amount of a cellulolytic enzyme composition in the presence of an effective amount of a polypeptide having cellulolytic enhancing activity of the present invention, wherein the presence of the polypeptide having cellulolytic enhancing activity increases the degradation of cellulose-containing material compared to the absence of the polypeptide having cellulolytic enhancing activity.

In another embodiment, the present invention relates to methods for producing a fermentation product, comprising: (a) saccharifying a cellulose-containing material with an effective amount of a cellulolytic enzyme composition in the presence of an effective amount of a polypeptide having cellulolytic enhancing activity of the present invention, wherein the presence of the polypeptide having cellulolytic enhancing activity increases the degradation of cellulose-containing material compared to the absence of the polypeptide having cellulolytic enhancing activity; (b) fermenting the saccharified cellulose-containing material of step (a) with one or more fermenting microorganisms to produce the fermentation product; and (c) recovering the fermentation product from the fermentation.

Table 1. List of target genes and reference to gene, transcript and protein sequences

<i>Gene ID in priority application</i>	<i>Gene ID herein</i>	Enzyme function	In priority application SEQ ID NO:			Herein SEQ ID NO:		
			Genomic seq	Coding seq	Amino acid seq	Genomic seq	Coding seq	Amino acid seq

Thiau2p7_006475	THIAU_1_01671	polysaccharide monooxygenase GH61	34	193	352	1	4	7
Thiau2p7_014916	THIAU_1_00455	cellobiohydrolase GH6	88	247	406	2	5	8
Thiau2p7_015044	THIAU_1_07449	polysaccharide monooxygenase GH61	90	249	408	3	6	9

Table 2. List of target genes and reference to exon boundaries

Gene ID	Genomic sequence (SEQ ID NO:)	Genomic sequence length	Exon boundaries (nucleotide positions) and exons
THIAU_1_01671	1	1135	1..246, 324..1007, 1082..1135
THIAU_1_00455	2	1714	1..92, 150..654, 815..1385, 1452..1714
THIAU_1_07449	3	789	1..394, 506..789

The present invention is illustrated in further details by the following non-limiting examples.

5

EXAMPLES

Example 1: Fermentation of the organism

Materials & Methods

In general, for each species, starter mycelium was grown in rich medium (either mycological broth or yeast malt broth (the latter being indicated with YM)) and then washed with water. The starter was then used to inoculate different liquid media or solid substrate and the resulting mycelium was used for RNA extraction and library construction.

Following are the medium recipes and the solid substrates with a referenced source (if available) as well as a table (**Table 3**) listing the media variations, since in some cases the basic recipes of the referenced source have been altered depending on the species grown. This is then followed by a summary of the specific species as grown in the examples.

A. Mycological broth

Per liter: 10 g soytone, 40 g D-glucose, 1 mL Trace Element solution, Double-distilled water;

Adjust pH to 5.0 with hydrochloric acid (HCl) and bring volume to 1 L with double-distilled water.

Trace Element Solution contains 2 mM Iron(II) sulphate heptahydrate ($\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$), 1 mM Copper (II) sulphate pentahydrate ($\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$), 5 mM Zinc sulphate heptahydrate

($\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$), 10 mM Manganese sulphate monohydrate ($\text{MnSO}_4 \cdot \text{H}_2\text{O}$), 5 mM Cobalt(II) chloride hexahydrate ($\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$), 0.5 mM Ammonium molybdate tetrahydrate ($(\text{NH}_4)_6\text{Mo}_7\text{O}_{24} \cdot 4\text{H}_2\text{O}$), and 95 mM Hydrochloric acid (HCl) dissolved in double-distilled water.

5 B. Yeast-Malt broth (YM)

(Reference: ATCC medium No. 200)

Per liter: 3 g yeast extract, 3 g malt extract, 5 g peptone, 10 g D-glucose, Double-distilled water to 1L.

10 C. Trametes Defined Medium (TDM)

(Reference: Reid and Place, Effect of Residual lignin type and amount on biological bleaching of kraft pulp by *Trametes versicolor*. *Applied Environmental Microbiology* 60: 1395-1400, 1994.)

Per liter: 10 g D-glucose, 0.75 g L-Asparagine monohydrate, 0.68 g Potassium phosphate monobasic (KH_2PO_4), 0.25 g Magnesium sulphate heptahydrate ($\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$), 15 mg Calcium chloride dihydrate ($\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$), 100 μg Thiamine hydrochloride, 1 ml Trace Element solution, 0.5 g TweenTM 80, Double distilled water;

Adjust pH to 5.5 with 3 M potassium hydroxide and bring volume to 1 L with double-distilled water.

20

Table 3. Variations of TDM media used for library construction

Variation	Description
TDM-1	Medium was prepared as in basic recipe described above.
TDM-2	Quantity of asparagine monohydrate was reduced to 0.15 g.
TDM-3	Manganese sulphate monohydrate was omitted from the medium.
TDM-4	The quantity of manganese sulphate monohydrate was raised to 0.2 mM final concentration in the medium.
TDM-5	The quantity of copper (II) sulphate pentahydrate was raised to 20 μM .
TDM-6	Glucose was replaced with 10 g per liter of cellulose (Solka-Floc, 200FCC)
TDM-7	Glucose was replaced with 10 g per liter of xylan from birchwood (Sigma Cat. # X-0502)
TDM-8	Glucose was replaced with 10 g per liter of wheat bran ¹ .
TDM-9	Glucose was replaced with 10 g per liter of citrus pectin (Sigma Cat. # P-9135).
TDM-10	Tween TM 80 was omitted from the medium.

TDM-11	The double-distilled water was replaced with whitewater ² collected from peroxide bleaching (which occurs during the manufacture of fine paper).
TDM-12	The double-distilled water was replaced with whitewater ² collected from newsprint manufacture.
TDM-13	Glucose was replaced with 5 g per liter of ground hardwood kraft pulp ³ .
TDM-14	The medium's pH was raised to 7.5.
TDM-15	The strain was incubated at 5°C above its optimum growth temperature.
TDM-16	The strain was incubated at 10°C below its optimum growth temperature.
TDM-17	One half of the double-distilled water was replaced with whitewater from newsprint manufacture. Glucose was omitted.
TDM-18	Potassium phosphate monobasic was replaced with 5 mM phytic acid from rice (Sigma Cat. # P3168).
TDM-19	Asparagine monohydrate was increased to 4 g per liter.
TDM-20	Asparagine monohydrate was increased to 4g per liter and glucose was replaced with 2% fructose.
TDM-21	Asparagine monohydrate was increased to 4 g per liter; 100 mL of double-distilled water was replaced with 100 mL kerosene ⁴ . Glucose was omitted.
TDM-22	Asparagine monohydrate was increased to 4 g per liter; 100 mL of double-distilled water was replaced with 100 mL hexadecane (Sigma cat. # H0255). Glucose was omitted.
TDM-23	Asparagine monohydrate was increased to 4 g per liter; one half of the double-distilled water was replaced with 25% whitewater from newsprint manufacture plus 25% white water from peroxide bleaching. Glucose was omitted.
TDM-24	Asparagine monohydrate was increased to 4 g per liter and the quantity of manganese sulphate monohydrate was raised to 0.2 mM final concentration in the medium.
TDM-25	Asparagine monohydrate was increased to 4 g per liter and manganese sulphate monohydrate was omitted from the medium.
TDM-26	Asparagine monohydrate was increased to 4 g per liter; and potassium phosphate monobasic was replaced with 5mM phytic acid from rice (Sigma Cat. # P3168).
TDM-27	Glucose was replaced with 10g per liter of olive oil (Sigma cat. # O1514)
TDM-28	One half of the double-distilled water was replaced with whitewater from peroxide bleaching. Glucose was omitted.
TDM-29	Glucose was replaced with 10 g per liter of tallow.
TDM-30	Glucose was replaced with 10 g per liter of yellow grease.
TDM-31	Glucose was replaced with 10 g per liter of defined lipid (Sigma cat. # L0288).
TDM-32	Glucose was replaced with 50 g per liter of D-xylose.
TDM-33	Glucose was replaced with 20 g per liter of glycerol and 20ml per liter of ethanol.

TDM-34	Glucose was reduced to 1 g per liter and 10 g per liter of bran was added.
TDM-35	Glucose was reduced to 1g per liter and 10 g per liter of pectin (Sigma Cat. # P-9135) was added.
TDM-36	Glucose was replaced with 10 g per liter of biodiesel.
TDM-37	Glucose was replaced with 10 g per liter of soy feedstock.
TDM-38	Glucose was replaced with 10g per liter of locust bean gum (Sigma cat # G0753).
TDM-39	One half of double-distilled water was replaced with a 1:1 ratio of whitewater from newsprint manufacture and white water from peroxide bleaching. Glucose was omitted.
TDM-40	The medium's pH was raised to 8.5.
TDM-41	One half of double-distilled water was replaced with whitewater from peroxide bleaching; plus yeast extract was added to 1 g per liter. Glucose was omitted.
TDM-42	Glucose was replaced with 5 g per liter of yellow grease and 5 g per liter of soy feedstock
TDM-43	Glucose was replaced with 20g per liter of fructose.
TDM-44	Glucose was replaced with 10 g per liter of cellulose (Solka-Floc, 200FCC) plus 1 g per liter of sophorose.
TDM-45	The medium's pH was raised to 8.84.

¹ Food grade wheat bran sourced from the supermarket was used.

² All whitewaters were sourced from Quebec paper mills by PAPRICAN on the Applicant's behalf.

³ Hardwood kraft pulp was sourced from Quebec paper mills by PAPRICAN on the Applicant's behalf.

⁴ Kerosene was sourced from a general hardware store.

D. Asparagine Salts Medium (AS):

(Reference: Ikeda et al., Laccase and Melanization in Clinically Important *Cryptococcus* Species Other Than *Cryptococcus neoformans* Journal of Clinical Microbiology 40: 1214-1218, 2002)

Per liter: 3.0 g D-glucose, 1.0 g L-Asparagine monohydrate, 3.0 g KH₂PO₄, 0.5 g Mg SO₄·7H₂O, 1 mg Thiamine.

Table 4. Variations of AS media used for library construction

Variation	Description
AS-1	Medium was prepared as in basic recipe described above.
AS-2	Glucose was replaced with 10 g per liter of pectin.
AS-3	One half of double-distilled water was replaced with a 1:1 ratio of whitewater from newsprint manufacture and white water from peroxide bleaching. Glucose was omitted.

E. Solid substrates used:

- | | |
|------|--|
| SS-1 | 5 g Wheat Bran. |
| SS-2 | 5 g Wheat bran plus 5 mL defined lipid. |
| SS-3 | 5 g Oat bran (food grade, sourced from supermarket). |

5 The *Thielavia australiensis* strain was grown according to the methods described above under the following growth conditions: TDM-1, -2, -3, -4, -5, -6, -7, -8, 9, -10, -13, -14, -15, -39; YM, whereby the following optimal growth temperature was used: 25°C.

The strains carrying the recombinant genes were grown according to the methods described above under the following growth conditions: minimal medium as described in Käfer et al., (1977, Adv. Genet. 19:33-131) except that the salt concentrations were raised ten-fold and the glucose concentration was 150 grams per liter, at 30°C.

Example 2: Genome sequencing and assembly

Genomic DNA was isolated from mycelium when the growth culture had reached the mid log phase. Genomic DNA was sequenced using the Roche 454 Titanium technology (http://www.454.com) to a genome coverage of over 20-fold according to the instructions of the manufacturer. The sequences were assembled using the Newbler and Celera assemblers (http://sourceforge.net/apps/mediawiki/wgs-assembler).

20 **Example 3: Building the cDNA library**

Total RNA was isolated from fungal cells or mycelia when the growth cultures had reached the late log phase. The mycelia were collected by filtration through Miracloth and washed with water by filtration. The mycelia were padded dry using paper towels, and frozen in liquid nitrogen and stored at -800°C. To extract total RNA, the frozen mycelia or cells were
25 ground to a fine powder in liquid nitrogen using pestle and mortar. Approximately 1-1.5 gram of frozen fungal powder was dissolved in 10 mL of TRIzol® reagent and RNA was extracted according to the manufacturer's protocol (Invitrogen Life Sciences, Cat. #15596-018). Following extraction, the RNA was dissolved at 1-1.5 mg/ml of DEPC-treated water.

The PolyATtract® mRNA Isolation Systems (Promega, Cat. #Z5300) was used to isolate poly(A)+RNA. In general, equal amounts of total RNA extracted from up to ten culture conditions were pooled. One milligram of total RNA was used for isolation of poly(A)+RNA according to the protocol provided by the manufacturer. The purified poly(A)+RNA was dissolved at 200-500 µg/mL of DEPC-treated water.

Five micrograms of poly(A)+RNA were used for the construction of cDNA library. Double-

stranded cDNA was synthesized using the ZAP-cDNA[®] Synthesis Kit (Stratagene, Cat. #200400) according to the manufacturer's protocol with the following modifications. An anchored oligo(dT) linker-primer was used in the first-strand synthesis reaction to force the primer to anneal to the beginning of the poly(A) tail of the mRNA. The anchored oligo(dT) linker-primer has the sequence:

5'-GAGAGAGAGAGAGAGAGAGAGAACTAGTCTCGAGTTTTTTTTTTTTTTTTTTVN-3' (SEQ ID NO: 10) where V is A, C, or G and N is A, C, G, or T. A second modification was made by adding trehalose at a final concentration of 0.6 M and betaine at a final concentration of 2 M in the buffer of the first-strand synthesis reaction to promote full-length synthesis. Following synthesis and size fractionation, fractions of double-stranded cDNA with sizes longer than 600 bp were pooled. The pooled cDNA was cloned directionally into the plasmid vector BlueScript KS+[®] (Stratagene) or a modified BlueScript KS+ vector that contained Gateway[®] (Invitrogen) recombination sites. The cDNA library was transformed into *E. coli* strain XL10-Gold ultracompetent cells (Stratagene, Cat. #Z00315) for propagation.

Bacterial cells carrying cDNA clones were grown on LB agar containing the antibiotic ampicillin for selection of plasmid-borne bacteria and X-gal and IPTG to use the blue/white system to screen for the presence cDNA inserts. The white bacterial colonies, those carrying cDNA inserts, were transferred by a colony-picking robot to 384-well MTP for replication and storage. Clones that were to be analyzed by sequencing were transferred to 96-well deep blocks using liquid-handling robots. The bacteria were cultured at 37°C with shaking at 150 rpm. After 24 hours of growth, plasmid DNA from the cDNA clones was prepared by alkaline lysis and sequenced from the 5' end using ABI 3730xl DNA analyzers (Applied Biosystems). The chromatograms obtained following single-pass sequencing of the cDNA clones were processed using Phred (available at <http://www.phrap.org>) to assign sequence quality values, Lucy as described in Chou and Holmes (2001, Bioinformatics, 17(12) 1093-1104) to remove vector and low quality sequences, and Phrap (available at <http://www.phrap.org/>) to assemble overlapping sequences derived from the same gene into contigs.

Example 4: Annotation

An in-house automated annotation pipeline was used to predict genes in the assembled genome sequence. The analysis pipeline used in part the ab initio tool Genemark[®] (<http://exon.biology.gatech.edu/>) for prediction. It also used the predictor Augustus (<http://augustus.gobics.de/>) trained on *de novo* assembled sequences and orthologous sequences for gene finding. Sequence similarity searches against the mycoCLAP[®]

(<http://cubique.fungalgenomics.ca/mycoCLAP/>) and NCBI non-redundant databases were performed with BLASTX as described in Altschul et al., (1997) (Nucleic Acids Res. 25(17): 3389-3402). Proteins encoding biomass-degrading enzymes possess conserved domains. We used the domains available at the European Bioinformatics Institute
5 (www.ebi.ac.uk/Tools/InterProScan/) to assist in the identification of target enzymes.

Proteins targeted to the extracellular space by the classical secretory pathway possess an N-terminal signal peptide, composed of a central hydrophobic core surrounded by N- and C-terminal hydrophilic regions. We used Phobius (available at <http://phobius.cgb.ki.se>) and SignalP[®] version 3 (available at <http://www.cbs.dtu.dk/services/SignalP>) to recognize the
10 presence of signal peptides encoded by the cDNA clones. The tools TargetP[®] (available at <http://www.cbs.dtu.dk/services/TargetP>) and Big-PI Fungal Predictor (available at http://mendel.imp.ac.at/gpi/fungi_server.html) were used to remove sequences that encode proteins which are targeted to the mitochondria or bound to the cell wall. Finally, sequences predicted to encode soluble secreted proteins by these automated tools were analyzed
15 manually. Clones that comprise full-length cDNAs which are predicted to encode soluble secreted proteins were sequenced completely. For genes identified from the genome sequence, oligonucleotide primers specific to the target genes were designed and used to PCR amplified the target genes from double-stranded cDNA or genomic DNA. The PCR amplified products were cloned into an appropriate expression vector for protein production in host cells. The
20 genomic, coding and polypeptides sequences were assigned SEQ ID NOs and functions, as summarized in **Table 1**.

Example 5: Assays for characterization of polypeptides

Polypeptides of the present invention may be additionally cloned into an expression vector,
25 expressed and characterized (e.g., in sugar release assays) for activity relating to their ability to breakdown and/or process biomass as described in WO/2012/92676, WO/2012/130950, and WO/2012/130964 using appropriate substrates (e.g., acid pre-treated corn stover, hot water treated washed wheat straw, or hot water treated washed corn fiber substrate). Soluble sugars that are released can be analyzed for example by proton NMR.

30 A number of assays may be used to characterize the polypeptides of the present invention. Selected non-limiting examples of such assays are described and/or referenced below. Of course, other assays not explicitly mentioned or referenced here may also be used, and the expression "can be" used below is intended to reflect this possibility. Furthermore, a person of skill in the art would be able to modify or adapt these and other assays, as necessary, to

characterize a particular polypeptide.

Acetylxyylan esterase CE5. Polypeptides of the present invention having this activity can be characterized as described in Water et al., Appl Environ Microbiol. (2012), 78(10): 3759-62; or Yang et al., International Journal of Molecular Sciences (2010), 11(12): 5143–5151.

5 **Adhesin.** Polypeptides of the present invention having this activity (reviewed in Dranginis et al., Microbiology and Molecular Biology Reviews (2007), 71(2): 282-294) can be characterized using techniques well known in the art (e.g. adhesion assays).

Aldose 1-epimerase (mutarotase, aldose mutarotase). Polypeptides of the present invention having this activity can be characterized as described in Timson and Reece, FEBS
10 Letters (2003), 543(1-3):21-24; and Villalobo et al., Exp. Parasitol. (2005) 110(3): 298-302.

Alpha-arabinofuranosidase. Polypeptides of the present invention having this activity can be characterized for example as described by Poutanen et al. (Appl. Microbiol. Biotechnol. 1988, 28, 425-432) using 5 mM p-nitrophenyl alpha-L-arabinofuranoside as substrates. The reactions may be carried out in 50 mM citrate buffer at pH 6.0, 40°C with a total
15 reaction time of 30 min. The reaction is stopped by adding 0.5 ml of 1 M sodium carbonate and the liberated p-nitrophenol is measured at 405 nm. Activity is expressed in U/ml. Furthermore, arabionofuranosidases may also be useful in animal feed compositions to increase digestibility. Corn arabinoxylan is heavily di-substituted with arabinose. In order to facilitate the xylan degradation it is advantageous to remove as many as possible of the arabinose substituents.
20 The in vitro degradation of arabinoxylans in a corn based diet supplemented with a polypeptide of the present invention having alpha-arabinofuranosidase activity and a commercial xylanase is studied in an in vitro digestion system, as described in WO/2006/114094.

Alpha-galactosidase. Polypeptides of the present invention having this activity can be characterized for example as described in US patent application publication No. US
25 2010/0273235 A1. Briefly, a synthetic substrate, 4-Nitrophenyl- α -D-galactoside is used and the release of p-Nitro-phenol is followed at a wavelength of 405 nm in a reaction buffer containing 100 mM sodium phosphate, 50 mM sodium chloride, pH 6.8 at 26°C.

Alpha-glucuronidase GH67. Polypeptides of the present invention having this activity can be characterized for example as described in Lee et al., J Ind Microbiol Biotechnol.
30 (2012), 39(8): 1245-51, or Nagy et al., J. Bacteriol. (2002), 184: 4925-4929.

Arabinogalactanase. Polypeptides of the present invention having this activity can be characterized for example as described in Yamamoto and Emi, Methods in Enzymology (1988), 160: 719-725.

Arabinoxylan arabinofuranohydrolase (AXH) GH43. Polypeptides of the present

invention having this activity can be characterized for example as described in Yoshida et al., Journal of Bacteriology (2010), 192(20): 5424–5436.

Arabinoxylan arabinofuranosidase GH62. Polypeptides of the present invention having this activity can be characterized for example as described in Sakamoto et al., Applied Microbiology and Biotechnology (2011), 90(1): 137-146.

Beta-galactosidase. Polypeptides of the present invention having this activity can be characterized for example using commercially available kits (e.g., β -Galactosidase Enzyme Assay System with Reporter Lysis Buffer, Cat. No. E2000, Promega).

Beta-glucanase. Polypeptides of the present invention having this activity can be characterized for example as described in US patent application publication number US 2012/0023626 A1; or in US patent No. 8,309,338.

Beta-glucosidase. Polypeptides of the present invention having this activity can be characterized for example as described in PCT application publication No. WO/2007/019442; or by using a commercially available kit (e.g., Beta-Glucosidase Assay Kit, Cat. No. KA1611, Abnova Corp).

Beta-mannanase. Polypeptides of the present invention having this activity can be characterized for example as described in European patent application No. EP 2261359 A1; or in PCT application publication No. WO2008009673A2.

Beta-xylosidase. Polypeptides of the present invention having this activity can be characterized for example as described in Wagschal et al., Applied and Environmental Microbiology (2005), 71(9): 5318–5323; or Shao et al., Appl Environ Microbiol. (2011), 77(3): 719–726.

Carbohydrate-binding cytochrome. Polypeptides of the present invention having this activity can be characterized for example as described in Yoshida et al., Applied and Environmental Microbiology (2005) 71(8): 4548–4555.

Cellobiohydrolase GH6. Polypeptides of the present invention having this activity can be characterized for example as described in Takahashi et al., Applied and Environmental Microbiology (2010), 76(19): 6583-6590.

Cellobiose dehydrogenase. Polypeptides of the present invention having this activity can be characterized for example as described in Schou et al., Biochem. J. (1998), 330: 565-571; or Baminger et al., J. Microbiol Methods. (1999), 35(3): 253-9.

Chitin deacetylase. Polypeptides of the present invention having this activity can be characterized for example as described in European patent application No. EP 0610320 B1.

Chitinase. Polypeptides of the present invention having this activity can be

characterized for example as described in US patent No. 7,087,810.

Chitotriosidase-1. Polypeptides of the present invention having this activity can be characterized for example as described in US patent No. 6,057,142.

5 **Cholinesterase.** Polypeptides of the present invention having this activity can be characterized for example as described in Abass Askar et al., Canadian Journal Veterinary Research (2011), 75(4): 261–270; or Cátia et al., PLoS One (2012), 7(3): e33975.

Cutinase. Polypeptides of the present invention having this activity can be characterized for example as described in US patent application publication No. US 2012/0028318 A1; or in Chen et al., J. Biol Chem. (2008), 283(38): 25854-62.

10 **Dehydrogenase.** Polypeptides of the present invention having this activity can be characterized for example as described in Mayer and Arnold, J. Biomol. Screen. (2002), 7(2): 135-140.

Endo-1,5-alpha-arabinanase. Polypeptides of the present invention having this activity can be characterized for example as described in US patent publication No. US 15 2012/0270263. More particularly, this assay of arabinase activity is based on colorimetrically determination by measuring the resulting increase in reducing groups using a 3,5-dinitrosalicylic acid reagent. Enzyme activity can be calculated from the relationship between the concentration of reducing groups, as arabinose equivalents, and absorbance at 540 nm. The assay is generally carried out at pH 3.5, but it can be performed at different pH values for the additional 20 characterization and specification of enzymes. Polypeptides of the present invention having this activity can also be characterized for example as described in Hong et al., Biotechnol Lett. (2009), 31(9): 1439-43.

Endo-1,6-beta-glucanase. Polypeptides of the present invention having this activity can be characterized for example as described in Bryant et al., Fungal Genet Biol. (2007), 25 44(8): 808-17; or in Oyama et al., Biosci Biotechnol Biochem. (2006), 70(7): 1773-5.

Endochitinase. Polypeptides of the present invention having this activity can be characterized for example as described in Wen et al., Biotechnol. Applied Biochem. (2002), 35: 213-219.

30 **Endoglucanase.** Polypeptides of the present invention having this activity can be characterized for example as described in US patent No. 8,063,267.

Exo-1,3-beta-galactanase GH43. Polypeptides of the present invention having this activity can be characterized for example as described in Ichinose et al., Appl Environ Microbiol. (2006), 72(5): 3515–3523.

Exo-1,3-beta-glucanase. Polypeptides of the present invention having this activity

can be characterized for example as described in O'Connell et al., Appl Microbiol Biotechnol. (2011), 89(3): 685-96; or Santos et al., J Bacteriol. (1979), 139(2): 333-338.

Exo-1,4-beta-xylosidase. Polypeptides of the present invention having this activity can be characterized for example as described in La Grange et al., Applied and Environmental Microbiology (2001), 67(12): 5512-5519.

Exoglucanase. Polypeptides of the present invention having this activity can be characterized for example as described in Creuzet et al., FEMS Microbiology Letters (1983), 20(3): 347-350; or Kruus et al., Journal of Bacteriology (1995), 177(6): 1641-1644.

Exo-glucosaminidase GH2. Polypeptides of the present invention having this activity can be characterized for example as described in Tanaka et al., Journal of Bacteriology (2003), 185(17): 5175-5181.

Exo-polygalacturonase. Polypeptides of the present invention having this activity can be characterized for example as described in Dong and Wang, BMC Biochem. (2011), 12: 51.

Expansin. Polypeptides of the present invention having this activity can be characterized for example as described in PCT application publication No. WO 2005/030965 A2; or in US patent No. 7,001,743.

Feruloyl esterase. Polypeptides of the present invention having this activity can be characterized for example as described in PCT application publication No. WO 2009/076122 A1.

Galactanase GH5. Polypeptides of the present invention having this activity can be characterized for example as described in Ichinose et al., Applied and Environmental Microbiology (2008), 74(8): 2379-2383.

Glucan 1,3-beta-glucosidase. Polypeptides of the present invention having this activity can be characterized for example as described in Boonvitthya et al., Biotechnol Lett (2012), 34(10): 1937-43.

Glycosidase. Polypeptides of the present invention having this activity can be characterized for example as described in US patent No. 8,119,383.

Hexosaminidase. Polypeptides of the present invention having this activity can be characterized for example as described in Wendeler and Sandhoff, Glycoconj J. (2009), 26(8):945-952.

Laccase. Polypeptides of the present invention having this activity can be characterized for example as described in Dedeyan et al., Appl Environ Microbiol. (2000), 66(3): 925-929.

Laminarinase GH55. Polypeptides of the present invention having this activity can be characterized for example as described in Ishida et al., J Biol Chem. (2009), 284(15): 10100–10109; or Kawai et al., Biotechnol Lett. (2006), 28(6): 365-71.

Licheninase (beta-D-glucan 4-glucanohydrolase). Polypeptides of the present invention having this activity can be characterized for example as described in Tang et al., J Agric Food Chem. (2012), 60(9): 2354-61.

Lipase. Polypeptides of the present invention having this activity can be characterized for example as described in US patent Nos. 7,662,602 and 7,893,232.

L-sorbose dehydrogenase. Polypeptides of the present invention having this activity can be characterized for example as described in Shinjoh et al., Applied and Environment Microbiology (1995), 61(2): 413-420.

Mixed-link glucanase. Polypeptides of the present invention having this activity can be characterized for example as described in Clark et al., Carbohydr Res. (1978), 61: 457–477.

Oxidoreductase. Polypeptides of the present invention having this activity can be characterized for example as described in Hommes et al., Anal Chem. (2013), 85(1): 283-291.

Para-nitrobenzyl esterase. Polypeptides of the present invention having this activity can be characterized for example as described in Moore and Arnold, Nat Biotechnol. (1996), 14(4): 458-67.

Pectate lyase. Polypeptides of the present invention having this activity can be characterized for example as described in Wang et al., BMC Biotechnology (2011), 11: 32.

Polysaccharide monooxygenase. Polypeptides of the present invention having this activity can be characterized for example as described in Kittl et al., Biotechnol Biofuels. (2012), 5(1):79.

Protease. Polypeptides of the present invention having this activity can be characterized for example as described in US patent application publication No. US 2005/0010037 A1.

Putative exoglucanase type C (1,4-beta-cellobiohydrolase; beta-glucancellobiohydrolase; exocellobiohydrolase I). Polypeptides of the present invention having this activity can be characterized for example as described in Dai et al., Applied Biochemistry and Biotechnology (1999), 79, Issue 1-3: 689-699.

Rhamnogalacturonan lyase PL4. Polypeptides of the present invention having this activity can be characterized for example as described in Mutter et al., Plant Physiol. (1998), 117: 153-163; or de Vries, Appl. Microbiol Biotechnol. (2003), 61: 10-20.

Swollenin. Polypeptides of the present invention having this activity can be

characterized for example as described in Jäger et al., *Biotechnol Biofuels*. (2011), 4: 33; or Saloheimo et al., *Eur J Biochem*. (2002), 269(17): 4202-11.

Tyrosinase. Polypeptides of the present invention having this activity can be characterized for example as described in US patent application publication No. US 2011/0311693 A1.

Xylan alpha-1,2-glucuronidase. Polypeptides of the present invention having this activity can be characterized for example as described in Ishihara, M. and Shimizu, K., "alpha-(1->2)-Glucuronidase in the enzymatic saccharification of hardwood xylan: Screening of alpha-glucuronidase producing fungi." *Journal Mokuzai Gakkaishi*, (1988) 34: 58-64.

Xylanase. Polypeptides of the present invention having this activity can be characterized for example as described in US patent application publication No. US 2012/0028306 A1; US patent No. 7,759,102; or PCT application publication No. WO 2006/078256 A2.

Xyloglucanase. Polypeptides of the present invention having this activity can be characterized for example as described in Master et al., *Biochem*. (2008), 411(1): 161-170.

Xylosidase/arabinosidase. Polypeptides of the present invention having this activity can be characterized for example as described in Whitehead and Cotta, *Curr Microbiol*. (2001), 43(4): 293-8; or Xiong et al., *Journal of Experimental Botany* (2007), 58(11): 2799-2810.

Although the present invention has been described hereinabove by way of specific embodiments thereof, it can be modified, without departing from the spirit and nature of the subject invention as defined in the appended claims.

Example 6. Identification of Thiau (*Thielavia australiensis*) genes that encode a secreted protein

General Molecular Biology Procedures:

Standard molecular cloning techniques such as DNA isolation, gel electrophoresis, enzymatic restriction modifications of nucleic acids, *E. coli* transformation etc. were performed as described by Sambrook *et al.*, 1989, (*Molecular cloning: a laboratory manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York and Innes et al. (1990) *PCR protocols, a guide to methods and applications*, Academic Press, San Diego, edited by Michael A. Innis et al). Primers were prepared by IDT (Integrated DNA Technologies). Sanger DNA sequencing was performed using an Applied Biosystem's 3730xl DNA Analyzer technology at the Innovation Centre (Génome Québec), McGill University in Montréal.

Construction of pGBFIN49 expression plasmids

Genes of interest were cloned into the expression vector pGBFIN-49. This vector is a derivative of pGBFIN-41 that contains the *A. niger glaA* promoter, *A. niger* TrpC terminator, *A. nidulans gpdA* promoter, gene encoding the pheomycin resistance gene, *A. niger glaA* terminator and an *E. coli* backbone. Figure 1 represents a schematic map of pGBFIN-49 and the complete nucleotide sequence is presented as SEQ ID NO: 11.

Details of the construction of pGBFIN-49 are as follows:

1. TtrpC terminator PCR amplification (0.7kb):

TtrpC terminator was PCR amplified using purified pGBFIN33 plasmid as a template. The following primers and PCR program were used:

Primer-3: 5'-GTCCGTCGCCGTCCTTCAccgccggtccgacg-3' (SEQ ID NO: 12)

Primer-4: 5'-GCGGCCGGCGTATTGGGTGttacggagc-3' (SEQ ID NO: 13)

Primer-4 is entirely specific to TtrpC 3' end. Primer-3 was designed to suit the LIC cloning strategy but also to keep TtrpC sequence as close as the original sequence. To do so, five adenines were replaced by thymines (underlined).

PCR master mix:

pGBFIN33	1 µl (5-10ng)
----------	---------------

Primer-3 (10mM)	1 µl
-----------------	------

Primer-4 (10mM)	1 µl
-----------------	------

dNTPs (2mM)	5 µl
-------------	------

HF Buffer (5x)	10 µl
----------------	-------

Phusion DNA pol	0.5 µl
-----------------	--------

Nuclease-free water	31.5 µl
---------------------	---------

Total	50 µl
-------	-------

PCR program: 1x98°C - 2 min; 25x (98°C - 30 sec, 68°C - 30 sec, 72°C - 1 min); 72°C - 7 min.

Reaction conditions: 5 µl of the PCR reaction was ran on 1.0% agarose gel and remaining was purified using QIAEX II gel Extraction kit (QIAGEN) and resuspended in nuclease-free water.

2. pGBFIN41 vector PCR amplification (8.3kb):

Vector backbone was PCR amplified using pGBFIN41 as a template. Primers were designed outside of the ccdA region (not included in pGBFIN49). The following primers and PCR program were used:

Primer-2: 5'-CACCCAATACGCCGGCCGCgcttcagacagctc-3' (SEQ ID NO: 14)

Primer-1C: 5'-GGTGTTTTGTTTGCTGGGGAAtgaagctcaggctctcagttgcgtc-3' (SEQ ID NO: 15)

Primer-2 contains a pgpdA-specific region and an extra sequence specific to TtrpC 3' end (also included in Primer-4). Primer-1C was designed to suit the LIC cloning strategy but also to keep PgalA region as close as the original sequence. To do so, three thymines were replaced by adenines (underlined).

PCR master mix:

pGBFIN41	1 µl (50 ng)
Primer-2 (10mM)	1 µl
Primer-1C (10mM)	1 µl
dNTPs (2mM)	5 µl
HF Buffer (5x)	10 µl
Phusion DNA pol	0.5 µl
DMSO	1 µl
<u>Nuclease-free water</u>	<u>30.5 µl</u>
Total	50 µl

PCR program: 1x98°C - 3 min; 10x (98°C - 30 sec, 68°C - 30 sec, 72°C - 5 min); 20 x (98°C - 30 sec, 68°C - 30 sec, 72°C - 5 min+10 sec/cycle); 72°C - 10 min.

Reaction conditions: 5 µl of the PCR reaction was ran on 0.5% agarose gel and remaining was purified using QIAEX II gel Extraction kit (QIAGEN) and resuspended in nuclease-free water.

3. pGBFIN41 + TtrpC overlap-extension PCR:

Overlap-extension / Long range PCR was performed to a) fused the two PCR pieces together; b) add SfoI restriction site to re-circulate the vector. No primers were used in the overlap-extension stage. Primer-11 and Primer-12 were used for the long range PCR reaction.

5 Primer-11: 5'-CACCGGCGCCGTCCGTCGCCGTCCTTC -3' (SEQ ID NO: 16)

Primer-12: 5'-ACGGCGCCGGTGTGTTTGTGCTGGGGATG -3' (SEQ ID NO: 17)

Primers-11 is specific to the LIC tag located on the TtrpC terminator while Primer-12 is specific to the LIC tag located on the PglA region. SfoI restriction site sequence is underlined.

10 A standard PCR master mix was prepared to perform overlap-extension PCR using pGBFIN41 and TtrpC purified PCR products as templates. No primers were added.

Overlap-extension master mix:

	TtrpC	1 µl
	pGBFIN41	9 µl
15	Buffer GC (5x)	10 µl
	dNTPs (2mM)	5 µl
	Phusion DNA pol	0.5 µl
	<u>Nuclease-free water</u>	<u>24.5 µl</u>
	Total	50 µl

20

PCR program – overlap (no primers): 1x 98°C - 2 min; 5x (98°C - 15 sec, 58° - C30 sec, 72°C - 5 min), 5x (98°C - 15 sec, 63°C - 30 sec, 72°C - 5 min), 5x (98°C - 15 sec, 68°C - 30 sec, 72°C - 5 min); 72°C - 10 min).

25 The overlap-extension PCR product was then, purified on QIAEX II column and 5 µl of the purified reaction was used as template DNA for Long range PCR step with primers-11 and -12.

PCR master mix:

	Overlap product	5ul
30	Primer-11 (10mM)	1 µl
	Primer-12 (10mM)	1 µl
	dNTPs (2mM)	5 µl
	HF Buffer (5x)	10 µl
	Phusion DNA pol	0.5 µl

DMSO	1 μ l
Nuclease-free water	26.5 μ l
Total	50 μ l

5 PCR program – Long range: 1x 98°C - 3 min; 10x (98°C - 30 sec, 68°C - 30 sec, 72°C - 5 min); 20 x (98°C - 30 sec, 68°C - 30 sec, 72°C - 5 min+10 sec/cycle); 72°C - 10 min.

Reaction conditions: 5 μ l of the PCR reaction was ran on 0.5% agarose gel and remaining was purified using QIAEX II gel Extraction kit and resuspended in nuclease-free water. Then, Sfol digestion was performed and digested product was purified using QIAEX II gel
10 extraction kit follow the procedure as described by the manufacture.

4. Ligation:

100 ng of the purified digested fragment was ligated to itself using 1 μ l of T4 DNA Ligase (New England Biolabs, M0202), and incubated at 16°C overnight. Enzyme inactivation was
15 performed at 65°C for 10 minutes.

Then, 10 μ l of ligation product were transformed in DH5 α E. coli competent cells and plated on 2xYT agar containing 100 ug/ml ampicillin. DNA extraction was performed on single colonies the next day. Restriction analysis and sequencing were done to confirm the
20 structure.

Cloning of *Thielavia australiensis* genes in *E. coli*

Cloning genes of interest in the pGBFIN-49 expression vector was performed using the Ligation-independent cloning (LIC) method according to Aslanidis, C., de Jong, P. (1990) Nucleic Acids Research Vol. 18 No. 20, 6069-6074.

25 Coding sequences from genes of interest were amplified by PCR using primers containing LIC tags which are homologous to *Pgla* and *TrpC* sequences in the pGBFIN-49 cloning vector fused to sequences homologous to the coding sequences of the gene of interest, and ei genomic DNA or cDNA as template. Primers have following sequences:

Forward primer: 5'-**CCCCAGCAACAAAACACCTCAGCAATG**...15-20 nucleotides
30 specific to each gene to be cloned (SEQ ID NO: 18)

Reverse primer: 5'- **GAAGGACGGCGACGGACTTCA**...15-20 nucleotides specific to each gene to be cloned (SEQ ID NO: 19)

PCR mix consists of following components:

Template (gDNA or cDNA) 1-10 ng/μl	1 μl
5X Phusion HF Buffer (Finnzymes)	10 μl
2 mM dNTPs	5 μl
LIC primer (F+R) mix 10 mM	0.5 μl
Phusion DNA Polymerase (Finnzymes)	0.5 μl
DMSO	1.5 μl
H ₂ O	31.5 μl
TOTAL	50 μl

PCR amplification was carried out with following conditions:

Cycle step	3-step protocol		Cycles
	Temp	Time	
Initial denaturation	98°C	30 s	1
Denaturation	98°C	10 s	10
Annealing	58°C	30 s	
Extension	72°C	30 s	
Denaturation	98°C	10 s	20
Annealing	68°C	30 s	
Extension	72°C	30 s	
Final extension	70°C	10 min	1
End of PCR storage	4°C	Hold	1

5

Following PCR, 90 μl milliQ water was added to each sample and the mix was purified using a MultiScreen PCR₉₆ Filter Plate (Millipore) according to manufacturer's instructions. The PCR product was eluted from the filter in 25 μl 10 mM Tris-HCl pH8.0.

Expression vector pGBFIN-49 was PCR amplified using primers with following sequences:

10

Forward primer: 5'- **GTCCGTCGCCGTCCTTCACCG** -3' (SEQ ID NO: 20)

Reverse primer: 5'- **GGTGTTTTGTTGCTGGGGATGAAGC** -3' (SEQ ID NO: 21)

(Primers are located at either site of the Sfol restriction site.)

PCR mix consists of following components:

pGBFIN-49 plasmid DNA (10ng/ μ l)	2 μ l
5X Phusion HF Buffer (Finnzymes)	20 μ l
2 mM dNTPs	10 μ l
LIC Primer mix (F+R) 10 mM	2 μ l
Phusion DNA Polymerase (Finnzymes)	1.5 μ l
DMSO	3 μ l
H ₂ O	61.5 μ l
TOTAL	100 μ l

5 PCR amplification was carried out with following conditions:

Cycle step	2-step PCR protocol		Cycles
	Temp.	Time	
Initial denaturation	98°C	2 min	1
Denaturation	98°C	10 s	35
Annealing+Extension	68°C	4 min +10s/cycle	
Final extension	70°C	10 min	1
End of PCR storage	4°C	Hold	1

Following PCR, 1 μ l DpnI was added to the PCR mix and digestion was allowed overnight at 37°C. Digested PCR product was purified using the Qiaquick PCR purification kit (Qiagen) according to manufacturer's instructions.

10 Obtained PCR fragments were treated with T4 DNA polymerase in the presence of dTTP to create single stranded tails at the ends of the PCR fragments. The single stranded tails of the PCR fragment are complementary to those at the vector, thus permitting non-covalent bi-molecular associations e.g. circularization between molecules.

15 Reaction mix of T4 DNA polymerase treatment pGBFIN-49 PCR fragment consists of following components:

Purified pGBFIN-49 PCR fragment	600 ng
---------------------------------	--------

10X Neb Buffer 2	2 μ l
25 mM dTTP	2 μ l
DTT 100 μ M	0.8 μ l
T4 DNA Polymerase 3U/ μ l	1 μ l
H ₂ O	Up to 20 μ l
TOTAL	20 μ l

Reaction mix of T4 DNA polymerase treatment of Gene of Interest (GOI) PCR fragment consists of following components:

Purified GOI PCR	5 μ l	5
10X NEB Buffer 2	2 μ l	
25 mM dATP	2 μ l	
DTT 100 μ M	0.8 μ l	
T4 DNA Polymerase 3U/ μ l	1 μ l	
H ₂ O	9.2 μ l	10
TOTAL	20 μ l	

Reaction conditions were as follows:

Step	Temperature °C	Duration
Annealing	22	30 min
Enzyme inactivation	75	20 min
End	4	Hold

Following T4 DNA polymerase treatment, 2 μ l pGBFIN-49 vector and 4 μ l of the GOI were mixed and incubated at room temperature allowing annealing of GOI fragment with pGBFIN-49 vector fragment. The bi-molecular forms are used to transform *E. coli*. Plasmid DNA of resulting transformants was isolated and verified by sequence analyses for correct amplification and cloning of the gene of interest.

20 **Transformation of *Thielavia australiensis* gene expression cassettes into *A. niger*.**

As host strain for enzyme production, *A. niger* GBA307 was used. Construction of *A. niger* GBA307 is described in WO2011009700.

Transformation of *A. niger* was performed essentially according to the method described by

Tilburn, J. et. al. (1983) *Gene* 26, 205-221 and Kelly, J & Hynes, M. (1985) *EMBO J.*, 4, 475-479 with the following modifications:

- Spores were grown for 16-24 hours at 30°C in a rotary shaker at 250 rpm in *Aspergillus* minimal medium. *Aspergillus* minimal medium contains per liter: 6 g NaNO₃; 0.52 g KCl; 1.52 g KH₂PO₄; 1.12 ml 4 M KOH; 0.52 g MgSO₄·7H₂O; 10 g glucose; 1 g casamino acids; 22 mg ZnSO₄·7H₂O; 11 mg H₃BO₃; 5 mg FeSO₄·7H₂O; 1.7 mg CoCl₂·6H₂O; 1.6 mg CuSO₄·5H₂O; 5 mg MnCl₂·2H₂O; 1.5 mg Na₂MoO₄·2 H₂O; 50 mg EDTA; 2 mg riboflavin; 2 mg thiamine-HCl; 2 mg nicotinamide; 1 mg pyridoxine-HCl; 0.2 mg panthotenic acid; 4 µg biotin; 10 ml Penicillin (5000IU/ml/Streptomycin (5000 UG/ml) solution (Invitrogen);
- Glucanex 200G (Novozymes) was used for the preparation of protoplasts;
- After protoplast formation (2-3 hours) 10 ml TB layer (per liter: 109.32 g Sorbitol; 100 ml 1 M Tris-HCl pH 7.5) was pipetted gently on top of the protoplast suspension. After centrifugation for 10 min at 4330 x g at 4°C in a swinging bucket rotor, the protoplasts on the interface were transferred to a fresh tube and washed with STC buffer (1.2 M Sorbitol, 10 mM Tris-HCl pH7.5, 50 mM CaCl₂). The protoplast suspension was centrifuged for 10 min at 1560 x g in a swinging bucket rotor and resuspended in STC-buffer at a concentration of 10⁸ protoplasts/ml;
- To 200 µl of the protoplast suspension, 20 µl ATA (0.4 M Aurintricarboxylic acid), the DNA dissolved in 10 µl in TE buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA), 100 µl of a PEG solution (20% PEG 4000 (Merck), 0.8M sorbitol, 10 mM Tris-HCl pH 7.5, 50 mM CaCl₂) was added;
- After incubation of the DNA-protoplast suspension for 10 min at room temperature, 1.5 ml PEG solution (60% PEG 4000 (Merck), 10 mM Tris-HCl pH7.5, 50 mM CaCl₂) was added slowly, with repeated mixing of the tubes. After incubation for 20 min at room temperature, suspensions were diluted with 5 ml 1.2 M sorbitol, mixed by inversion and centrifuged for 10 min at 2770 x g at room temperature.
- The protoplasts were resuspended gently in 1 ml 1.2 M sorbitol and plated onto selective regeneration medium consisting of *Aspergillus* minimal medium without riboflavin, thiamine.HCl, nicotinamide, pyridoxine, panthotenic acid, biotin, casamino acids and glucose, supplemented with 150 µg/ml Phleomycin (Invitrogen), 0.07 M NaNO₃, 1 M sucrose, solidified with 2 % bacteriological agar #1 (Oxoid, England). After incubation for 5-10 days at 30°C, single transformants were isolated on PDA (Potato Dextrose Agar (Difco) supplemented with 150 µg/ml Phleomycin in 96 wells MTP. After 5-7 days growth at 30°C single transformants were used for MTP fermentation.

Aspergillus niger shake flask fermentation

Approximately 1×10^6 - 1×10^7 spores were inoculated in 20 ml pre-culture medium containing
5 Maltose 30 g/l; Peptone (aus casein) 10 g/l; Yeast extract 5 g/l; KH_2PO_4 1 g/l; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$
0.5 g/l; ZnCl_2 0.03 g/l; CaCl_2 0.02 g/l; $\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$ 0.01 g/l; $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 0.3 g/l; Tween-80 3
g/l; pH 5.5. After growing overnight at 34°C in a rotary shaker, 10-15 ml of the growing culture
was inoculated in 100 ml main culture containing Glucose $\cdot \text{H}_2\text{O}$ 70 g/l; Peptone (aus casein) 25
g/l; Yeast extract 12.5g/l; K_2SO_4 2 g/l; KH_2PO_4 1g/l; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 0.5 g/l; ZnCl_2 0.03 g/l;
10 CaCl_2 0.02 g/l; $\text{MnSO}_4 \cdot 1\text{H}_2\text{O}$ 0.009 g/l; $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ 0.003 g/l; pH 5.6.

Note: for GH61 enzymes the culture media were supplemented with 10 μM CuSO_4 .

Main cultures were grown until all glucose was consumed as measured with Combur Test N
strips (Roche) which was the case mostly after 4-7 days of growth. Culture supernatants were
harvested by centrifugation for 10 minutes at 5000 x g followed by germ-free filtration of the
15 supernatant over 0.2 μm PES filters (Nalgene).

Shake flask concentration and protein concentration determination with TCA-biuret method

In order to obtain greater amounts of material for further testing the fermentation supernatants
20 obtained as described above (volume between 75 and 100 ml) were concentrated using a 10
kDa spin filter to a volume of approximately 5 ml. Subsequently, the protein concentration in the
concentrated supernatant was determined via a TCA-biuret method.

Concentrated protein samples (supernatants) were diluted with water to a concentration
between 2 and 8 mg/ml. Bovine serum albumin (BSA) dilutions (0, 1, 2, 5, 8 and 10 mg/ml)
25 were made and included as samples to generate a calibration curve. Of each diluted protein
sample 270 μl was transferred into a 10 ml tube containing 830 μl of a 12% (w/v) trichloro acetic
acid solution in acetone and mixed thoroughly. Subsequently, the tubes were incubated on ice
water for one hour and centrifuged for 30 minutes, at 4°C and 6000 rpm. The supernatant was
discarded and pellets were dried by inverting the tubes on a tissue and letting them stand for 30
30 minutes at room temperature. Next, 3 ml BioQuant Biuret reagent mix was added to the pellet in
the tube and the pellet was solubilized upon mixing followed by addition of 1 ml water. The tube
was mixed thoroughly and incubated at room temperature for 30 minutes. The absorption of the
mixture was measured at 546 nm with a water sample used as a blank measurement and the
protein concentration was calculated via the BSA calibration line.

Enzyme activity assays

A. niger strains expressing *Thielavia australiensis* clones were grown in shake flask, as described above, in order to obtain greater amounts of material for further testing. The fermentation supernatants (volume between 40 and 80ml) were concentrated using a 10 kDa spin filter to a volume of approximately 5 ml. Subsequently, the protein concentration in the concentrated supernatant was determined via a TCA-biuret method, as described above.

The (hemi-)cellulase activity of these protein samples was tested in an assay where the supernatants were spiked on top of an enzyme base mix in the presence of 10% (w/w) acid pretreated corn stover. 'To spike' or 'spiking of' a supernatant or an enzyme indicates in this context the addition of a supernatant or an enzyme to a (hemi-)cellulase base mix. The feedstock solution was prepared via the dilution of a concentrated feedstock solution with water. Subsequently the pH was adjusted to pH 4.5 with a 4M NaOH solution. The proteins were spiked based on dosage in a total volume of 20 ml at a feedstock concentration of 10% aCS (w/w) in a 40-ml centrifuge bottle (Nalgene Oakridge). All experiments were performed at least in duplicate and were incubated for 72 hours at 65°C in an oven incubator (Techne HB-1D hybridization oven) while rotating at set-point 3. After incubation, the samples were centrifuged and soluble sugars were analysed by HPLC as described below.

Soluble sugar analysis by HPLC

The sugar content of the samples after enzymatic hydrolysis were analyzed using a High-Performance Liquid Chromatography System (Agilent 1100) equipped with a refraction index detector (Agilent 1260 Infinity). The separation of the sugars was achieved by using a 300 X 7.8 mm Aminex HPX-87P (Bio rad cat no 125-0098) column; Pre-column: Micro guard Carbo-P (Bio Rad cat no 125-0119); mobile phase was HPLC grade water ; flow rate of 0.6 ml/min and a column temperature of 85 °C. The injection volume was 10 µl.

The samples were diluted with HPLC grade water to a maximum of 10 g/l glucose and filtered by using 0.2 µm filter (Afridisc LC25mm syringe filter PVDF membrane) . The glucose was identified and quantified according to the retention time, which was compared to the external glucose standard (D-(+)-Glucose Sigma cat no: G7528) ranging from 0.2; 0.4; 1.0; 2.0 g/l.

Identification of THIAU (*Thielavia australiensis*) genes that encode a secreted protein

Genes were identified that based on curation (described above) encoded a secreted protein. A list of these genes is shown in Table 2.

Example 7. Improvement of a thermophilic cellulase mixture composed of three enzymes by a *Thielavia australiensis* protein in an activity assay.

5 The cellulase enhancing activity of two *Thielavia australiensis* GH61 proteins was further analysed. The supernatants of the *A. niger* expressing GH61 shake flask fermentations were concentrated and spiked in a dosage of 1.8 mg/gDM on top of a base activity of a three enzyme base mix (3.2 mg/gDM composed of: BG at 0.45 g/gDM, CBHI at 1.5 mg/gDM and CBHII at 1.25 mg/gDM) at a feedstock concentration of 10% (w/w) aCS, as described above. As a
10 negative control, the 3 enzyme base mix was also tested. All experiments were performed at least in duplicate and were incubated for 72 hours at 65°C in an oven incubator (Techne HB-1D hybridization oven) while rotating at set-point 3. After incubation, the samples were centrifuged and soluble sugars were analysed by HPLC as described above.

Addition of this *Thielavia australiensis* GH61 proteins showed increased sugar release as
15 shown below in Table 5.

Protein ID	glucose (g/l)
THIAU_1_01671	23.7
THIAU_1_07449	28.5
3 enzyme mix no GH61	16.7

Table 5: Effect of *Thielavia australiensis* GH61 proteins spiked on top of a 3E mix using aCS substrate.

20 In another experiment, the cellulase enhancing activity of *Thielavia australiensis* CBHII protein (THIAU_1_00455) was further analysed. The THIAU_1_00455 gene was cloned and expressed in *A. niger* as described above. The supernatant of an *A. niger* expressing THIAU_1_00455 shake flask fermentation was concentrated and spiked in a dosage of 1.5 mg/gDM on top of a
25 base activity of a three enzyme base mix (3.5 mg/gDM composed of: BG at 0.45 g/gDM, CBHI at 1.25 mg/gDM and GH61 at 1.8 mg/gDM) at a feedstock concentration of 10% (w/w) aCS, as described above. As a negative control, the 3 enzyme base mix was also tested. All experiments were performed at least in duplicate and were incubated for 72 hours at 65°C in an oven incubator (Techne HB-1D hybridization oven) while rotating at set-point 3. After incubation,
30 the samples were centrifuged and soluble sugars were analysed by HPLC as described below.

Addition of this *Thielavia australiensis* CBHII protein showed increased sugar release as shown below in Table 6.

Protein ID	glucose (g/l)
THIAU_1_00455	20.7
3 enzyme mix no CBHII	17.2

- 5 Table 6: Effect of CBHII THIAU_1_00455 protein spiked on top of a 3E mix using aCS substrate.

CLAIMS:

1. A process for degrading biomass or pretreated biomass to sugars wherein an enzyme is used comprising
 - a. a polypeptide comprising an amino acid sequence of any one of SEQ ID NOs: 7-9;
 - b. a polypeptide comprising an amino acid sequence that is at least 85%, preferably at least 90%, more preferably at least 95%, even more preferably at least 96%, 97%, 98% or 99% amino acid sequence identity to the polypeptide defined in (a);
 - c. a polypeptide comprising an amino acid sequence encoded by the nucleic acid sequence of any one of SEQ ID NOs: 1-6;
 - d. a polypeptide comprising an amino acid sequence encoded by a polynucleotide molecule that hybridizes under medium-high stringency conditions, high stringency conditions, or very high stringency conditions with the full-length complement of a polynucleotide molecule comprising the nucleic acid sequence of any one of SEQ ID NOs: 1-6;
 - e. a polypeptide comprising an amino acid sequence encoded by a polynucleotide molecule having 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%, or at least 99% nucleic acid sequence identity to a polynucleotide comprising the nucleic acid sequence of any one of SEQ ID NOs: 1-6; or
 - f. a polypeptide comprising an amino acid sequence encoded by any one of the exonic nucleic acid sequences corresponding to positions: 1-92, 150-654, 815-1385, 1452-1714 of SEQ ID NO: 2; 1-246, 324-1007, 1082-1135 of SEQ ID NO: 1; 1-394, 506-789 of SEQ ID NO: 3.
2. A process for degrading biomass or pretreated biomass to sugars according to claim 1 wherein the enzyme is a cellobiohydrolase, GH6, or polysaccharide monooxygenase or GH61 or an activity according to Table 1.
3. A process for degrading biomass or pretreated biomass to sugars according to claim 1 or 2 wherein the polypeptide is obtainable from *Thielavia australiensis*.
4. A process for degrading biomass or pretreated biomass to sugars according to any one of claims 1 to 3 wherein the formed sugars are converted into ethanol.

5. A process for degrading biomass or pretreated biomass to sugars according to any one of claims 1 to 4 further comprising adding a cellulase or cellulases.

6. A process for degrading biomass or pretreated biomass to sugars according to any one of claims 1 to 5 wherein the cellulolytic material or lignin is pretreated.

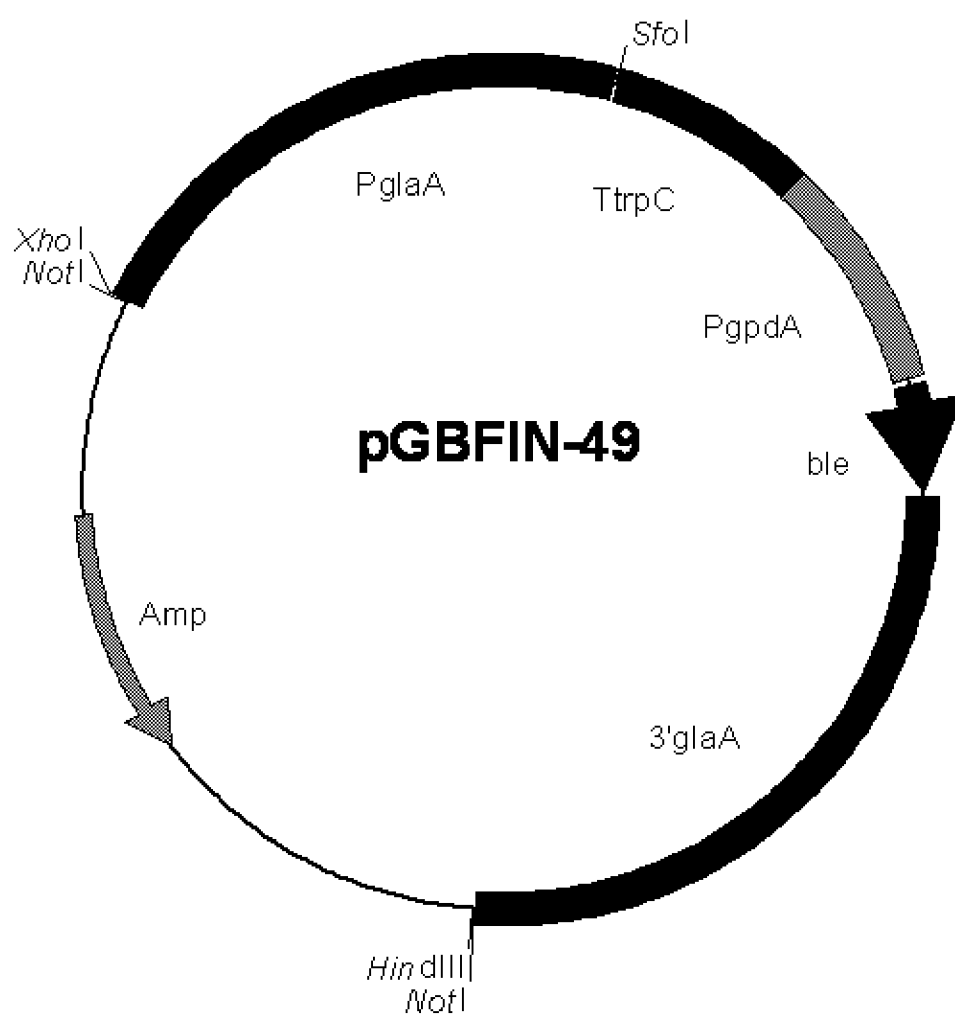


Fig. 1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2014/054954

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:
- a. (means)
- ☐ on paper
- ☒ in electronic form
- b. (time)
- ☒ in the international application as filed
- ☐ together with the international application in electronic form
- ☐ subsequently to this Authority for the purpose of search
2. ☐ In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2014/054954

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-6(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/054954

A. CLASSIFICATION OF SUBJECT MATTER INV. C12P19/02 C12P19/14 C13K1/02 C12N9/42 C12P7/10 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12P C13K C12N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data, BIOSIS, COMPENDEX, EMBASE, FSTA, Sequence Search		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2005/191736 A1 (BROWN KIMBERLY [US] ET AL) 1 September 2005 (2005-09-01) cited in the application abstract; claims 8, 48; figure 35; examples 1-28; sequences 1-10 paragraphs [0250] - [0255], [0286] - [0287]	1-6
X	US 2010/159494 A1 (SWEENEY MATTHEW DAVID [US] ET AL) 24 June 2010 (2010-06-24) abstract; claims 1-10; example 15; sequences 1, 2 paragraphs [0007], [0109], [0142], [0149]	1-6
----- ----- ----- ---/---		
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
16 June 2014		04/09/2014
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Schröder, Gunnar

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2014/054954

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2012/024698 A1 (CODEXIS INC [US]; BAIDYARROY DIPNATH [US]; CLARK LOUIS [US]; CAMPOPIANO) 23 February 2012 (2012-02-23) abstract; claims 32-50; figures 1-3; sequences 20, 49 -----	1,2,4-6
X	WO 2012/125925 A2 (DANISCO US INC [US]; MITCHINSON COLIN [US]; LI MIAN [US]; KELEMEN BRAD) 20 September 2012 (2012-09-20) abstract; claim 1 paragraph [0070]; sequence 17 -----	1,2,4-6
X	WO 2009/085859 A2 (NOVOZYMES AS [DK]; HARRIS PAUL [US]; MAIYURAN SUCHINDRA [US]; BROWN KI) 9 July 2009 (2009-07-09) abstract; claims 1, 17-20; sequences 1, 2 page 59, lines 10-23 -----	1,2,4-6

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2014/054954

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2005191736 A1	01-09-2005	CN 1980953 A	13-06-2007
		DK 1713825 T3	05-11-2012
		DK 2301958 T3	16-06-2014
		DK 2305702 T3	16-06-2014
		DK 2305703 T3	16-06-2014
		DK 2314605 T3	16-06-2014
		EP 1713825 A2	25-10-2006
		EP 2301958 A1	30-03-2011
		EP 2305702 A1	06-04-2011
		EP 2305703 A1	06-04-2011
		EP 2308890 A1	13-04-2011
		EP 2314605 A1	27-04-2011
		ES 2391826 T3	30-11-2012
		ES 2468366 T3	16-06-2014
		ES 2469840 T3	20-06-2014
		ES 2469874 T3	20-06-2014
		ES 2472669 T3	02-07-2014
		US 2005191736 A1	01-09-2005
		US 2008206815 A1	28-08-2008
		US 2010197556 A1	05-08-2010
		US 2014093940 A1	03-04-2014
		WO 2005074647 A2	18-08-2005
US 2010159494 A1	24-06-2010	US 2010159494 A1	24-06-2010
		WO 2010080527 A1	15-07-2010
WO 2012024698 A1	23-02-2012	CA 2807702 A1	22-02-2012
		CN 103080306 A	01-05-2013
		EP 2606131 A1	26-06-2013
		US 2012083019 A1	05-04-2012
		US 2014186896 A1	03-07-2014
		WO 2012024698 A1	23-02-2012
WO 2012125925 A2	20-09-2012	CA 2830508 A1	20-09-2012
		CN 103842515 A	04-06-2014
		EP 2686434 A2	22-01-2014
		JP 2014511673 A	19-05-2014
		KR 20140017612 A	11-02-2014
		SG 192098 A1	30-08-2013
		US 2014134677 A1	15-05-2014
		WO 2012125925 A2	20-09-2012
WO 2009085859 A2	09-07-2009	CA 2709371 A1	09-07-2009
		CN 101970471 A	09-02-2011
		EP 2245050 A2	03-11-2010
		US 2010306881 A1	02-12-2010
		WO 2009085859 A2	09-07-2009

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-6(partially)

Process for degrading biomass or pretreated biomass to sugars using an enzyme comprising an amino acid sequence of SEQ ID No: 7 or sequences having at least 85% identity therewith, or an amino acid sequence encoded by the nucleic acid sequence of SEQ ID NO: 1 or 4 or sequences having at least 85% identity therewith

2. claims: 1-6(partially)

Process for degrading biomass or pretreated biomass to sugars using an enzyme comprising an amino acid sequence of SEQ ID No: 8 or sequences having at least 85% identity therewith, or an amino acid sequence encoded by the nucleic acid sequence of SEQ ID NO: 2 or 5 or sequences having at least 85% identity therewith

3. claims: 1-6(partially)

Process for degrading biomass or pretreated biomass to sugars using an enzyme comprising an amino acid sequence of SEQ ID No: 9 or sequences having at least 85% identity therewith, or an amino acid sequence encoded by the nucleic acid sequence of SEQ ID NO: 3 or 6 or sequences having at least 85% identity therewith
