US 20230203461A1

(19) **United States**

(12) **Patent Application Publication**

Bruce et al.

(10) **Pub. No.: US 2023/0203461 A1**

(43) **Pub. Date: Jun. 29, 2023**

(54) **BETA-ETHERASES FOR LIGNIN DEPOLYMERISATION**

(71) Applicants: **The University of York**, York (GB); **Wisconsin Alumni Research Foundation**, Madison, WI (US)

(72) Inventors: **Neil Bruce**, York (GB); **Nicola Oates**, York (GB); **John Ralph**, Madison, WI (US)

(73) Assignees: **The University of York**, York (GB); **Wisconsin Alumni Research Foundation**, Madison, WI (US)

(21) Appl. No.: **17/791,144**

(22) PCT Filed: **Jan. 9, 2021**

(86) PCT No.: **PCT/EP2021/050317**

§ 371 (c)(1),
(2) Date: **Jul. 6, 2022**

(30) **Foreign Application Priority Data**

Jan. 10, 2020 (GB) 2000378.6

Publication Classification

(51) **Int. Cl.**
C12N 9/14 (2006.01)
C12N 15/80 (2006.01)

(52) **U.S. Cl.**
CPC ***C12N 9/14*** (2013.01); ***C12N 15/80*** (2013.01); ***C12Y 303/00*** (2013.01)

(57) **ABSTRACT**

The present application relates to nucleic acids encoding polypeptides with β-etherase activity; polypeptides with β-etherase activity; vectors comprising said nucleic acids for the production of recombinant β-etherase; cells, for example microbial cells transformed with nucleic acids encoding β-etherase activity and vectors, including nucleic acids encoding β-etherases; a composition comprising β-etherases suitable for processing lignocellulose and a method that uses β-etherases or compositions comprising β-etherases in the processing of lignocellulose and related polysaccharides.

Specification includes a Sequence Listing.

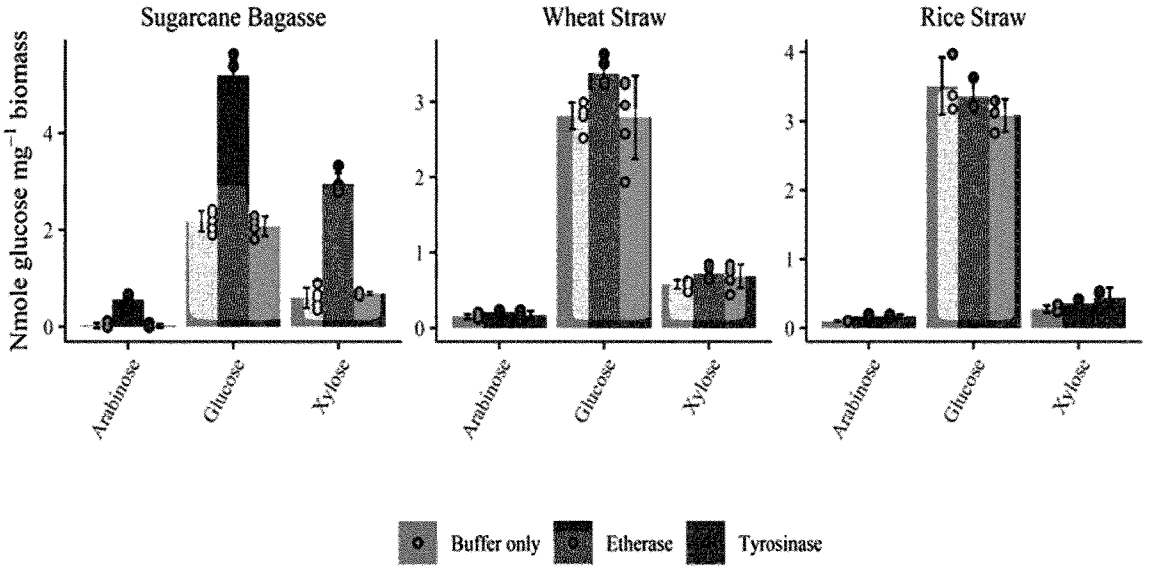


FIG. 1A

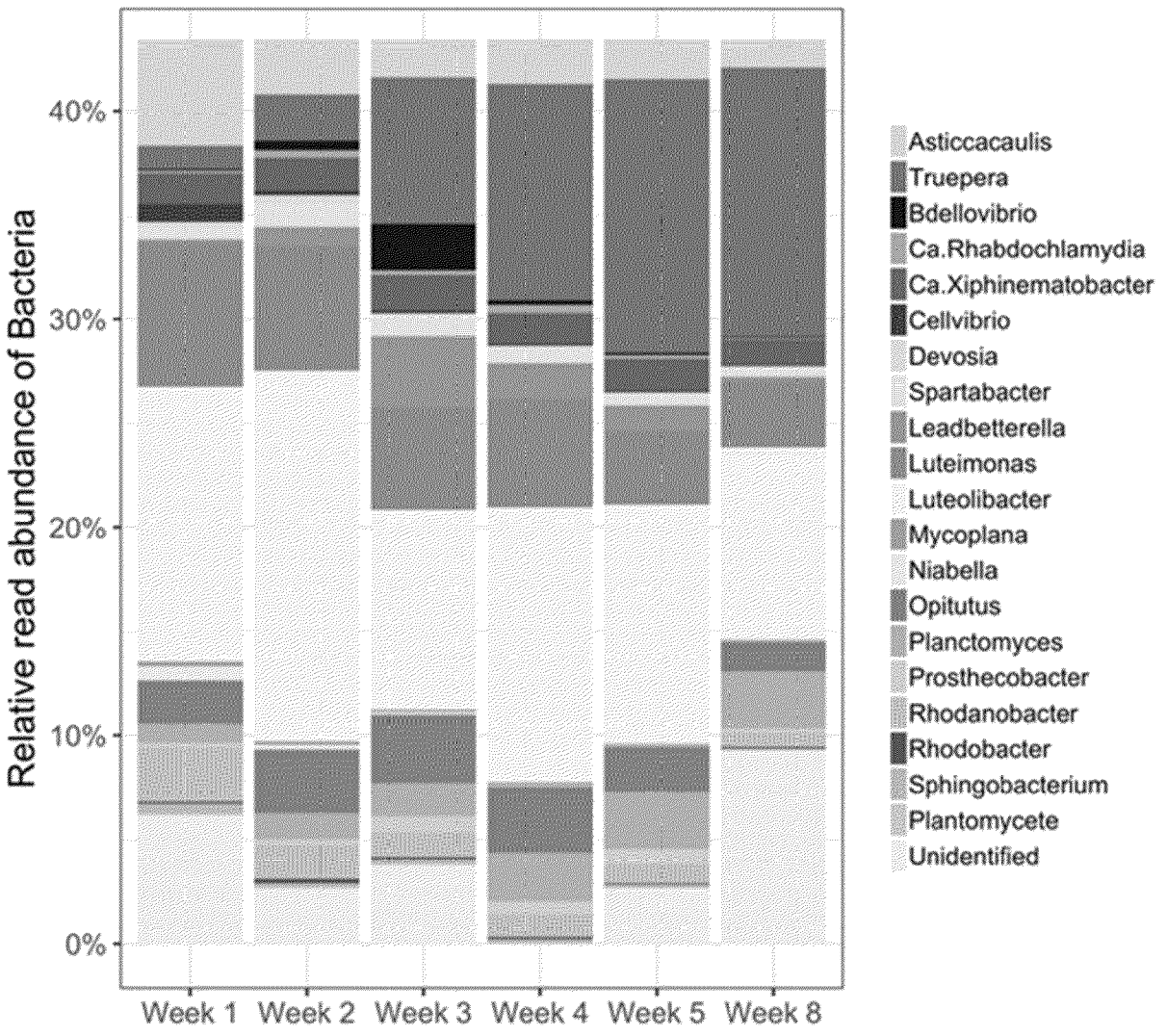

Relative read abundance of Bacteria
40%
30%
20%
10%
0%
Week 1
Week 2
Week 3
Week 4
Week 5
Week 8
Asticcacaulis
Truepera
Bdellovibrio
Ca.Rhabdochlamydia
Ca.Xiphinematobacter
Cellvibrio
Devosia
Spartabacter
Leadbetterella
Luteimonas
Luteolibacter
Mycoplana
Niabella
Opitutus
Planctomyces
Prosthecobacter
Rhodanobacter
Rhodobacter
Sphingobacterium
Plantomycete
Unidentified

FIG. 1B

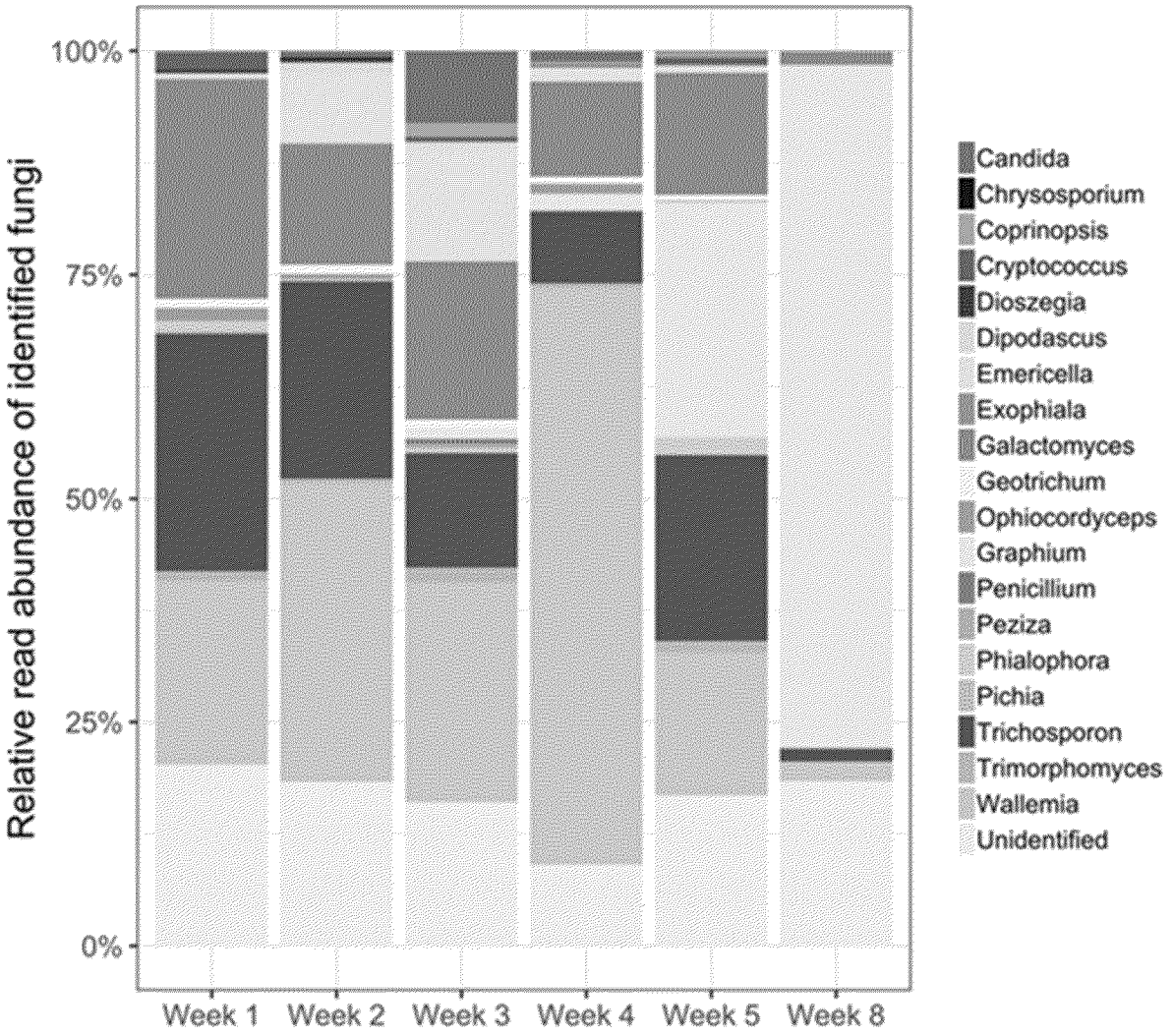

Relative read abundance of identified fungi
100%
75%
50%
25%
0%
Week 1
Week 2
Week 3
Week 4
Week 5
Week 8
Candida
Chrysosporium
Coprinopsis
Cryptococcus
Dioszegia
Dipodascus
Emericella
Exophiala
Galactomyces
Geotrichum
Ophiocordyceps
Graphium
Penicillium
Peziza
Phialophora
Pichia
Trichosporon
Trimorphomyces
Wallemia
Unidentified

FIG. 2A

FIG. 2B

FIG. 2C

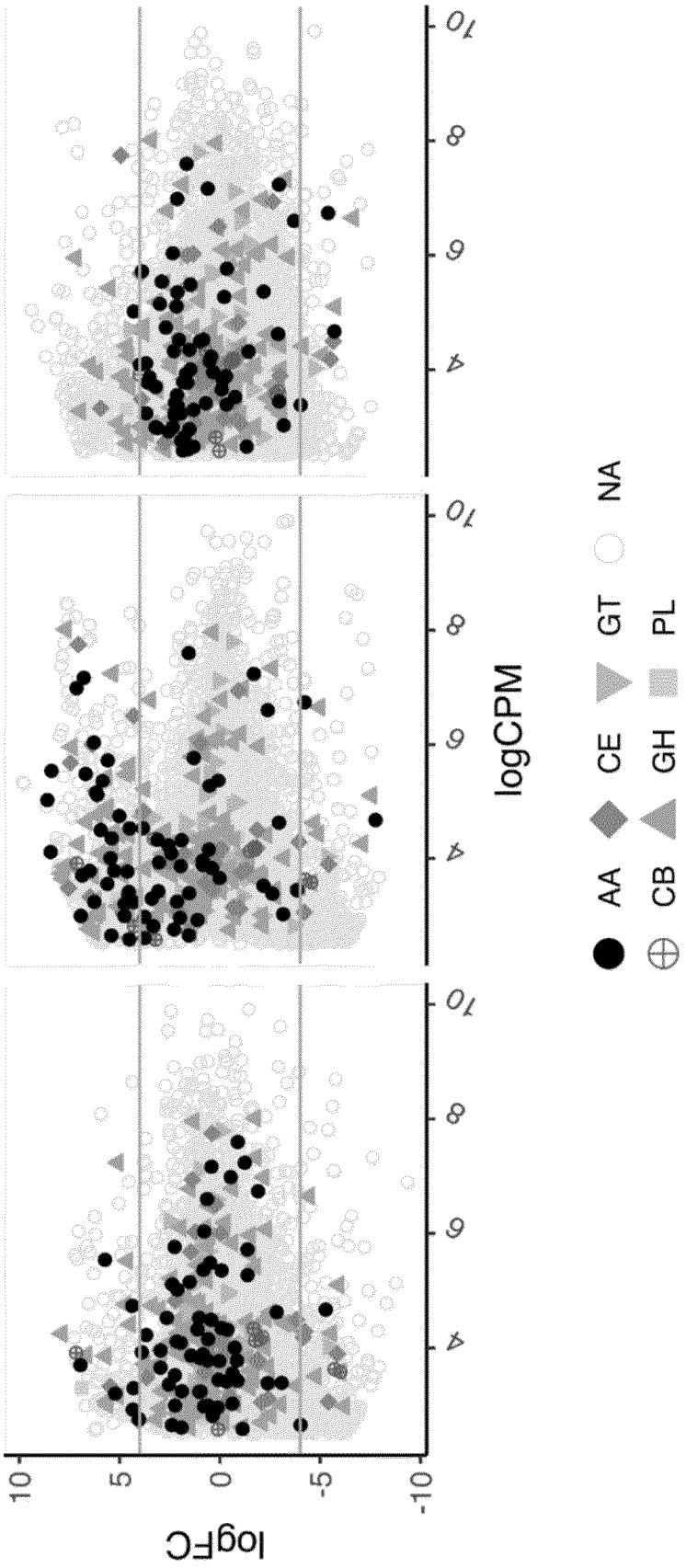


FIG. 3

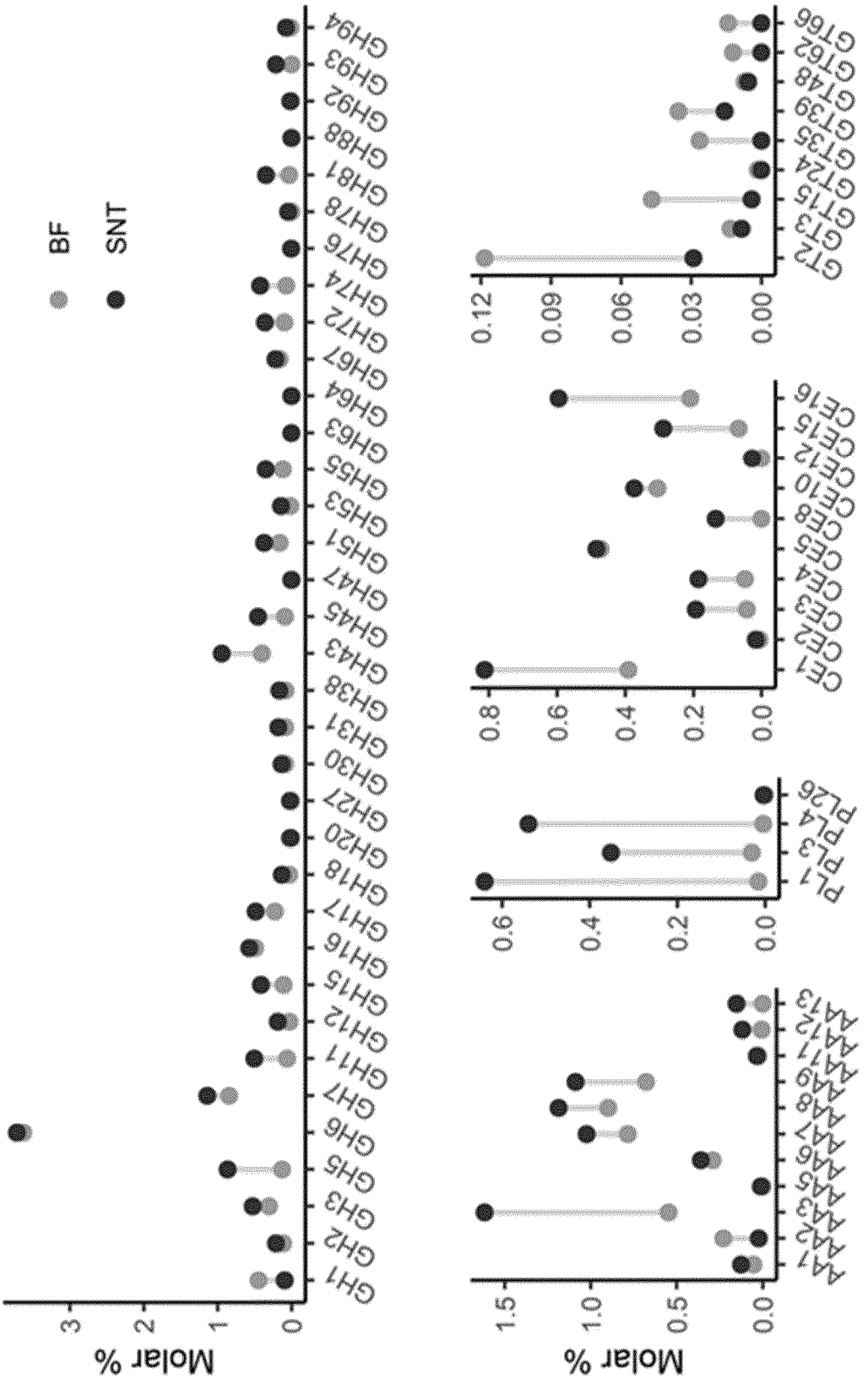
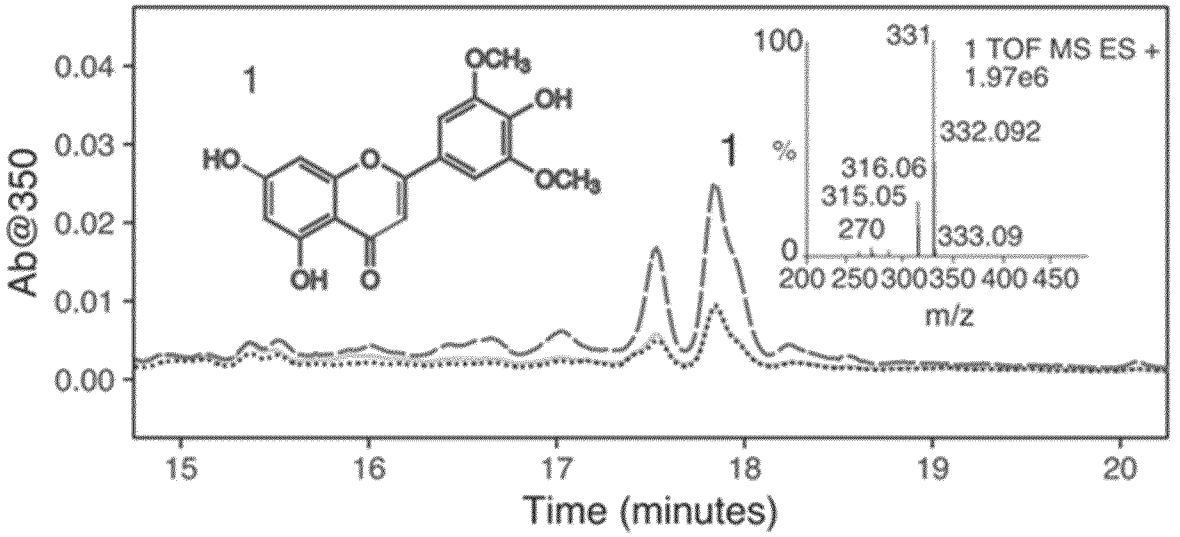

BF
SNT
Molar %
GH1
GH2
GH3
GH5
GH6
GH7
GH11
GH12
GH15
GH16
GH17
GH18
GH20
GH27
GH30
GH31
GH38
GH43
GH45
GH47
GH51
GH53
GH55
GH63
GH64
GH67
GH72
GH74
GH76
GH78
GH81
GH88
GH92
GH93
GH94
AA1
AA2
AA3
AA5
AA6
AA7
AA8
AA9
AA11
AA12
AA13
PL1
PL3
PL4
PL26
CE1
CE2
CE3
CE4
CE5
CE8
CE10
CE12
CE15
CE16
GT2
GT3
GT15
GT24
GT35
GT39
GT48
GT62
GT66

FIG. 4A


Ab@350
0.04
0.03
0.02
0.01
0.00
15
16
17
18
19
20
Time (minutes)
1
1
100
%
0
331
332.092
316.06
315.05
270
333.09
200 250 300 350 400 450
m/z
1 TOF MS ES +
1.97e6

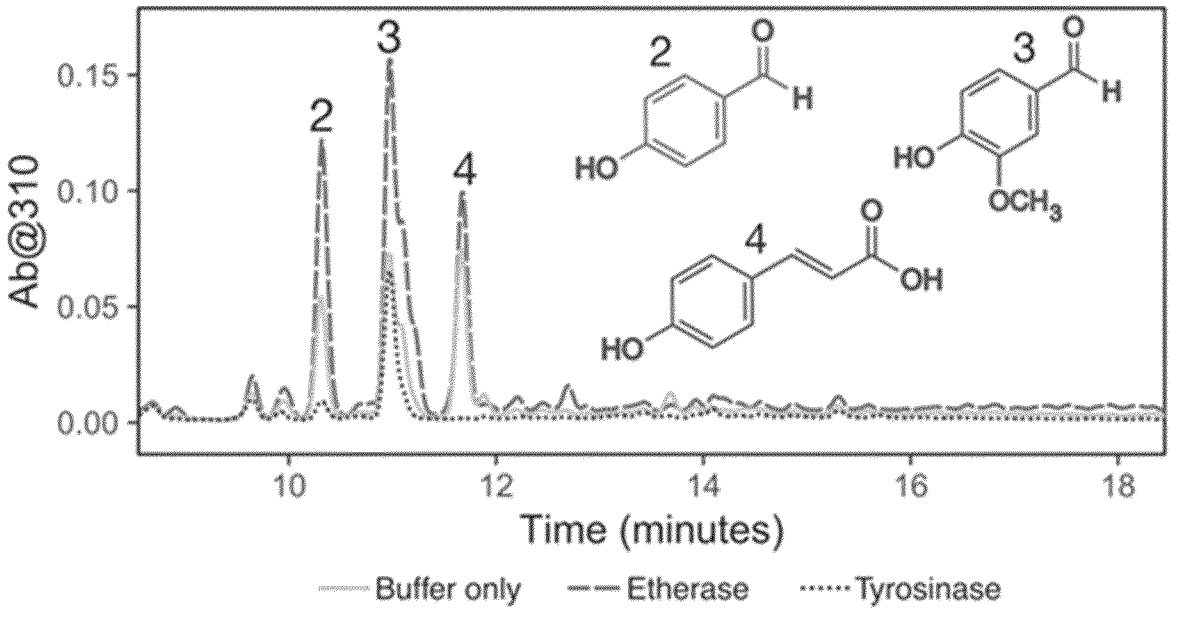

Ab@310
0.15
0.10
0.05
0.00
10
12
14
16
18
Time (minutes)
2
3
4
Buffer only
Etherase
Tyrosinase

FIG. 4B

FIG. 5

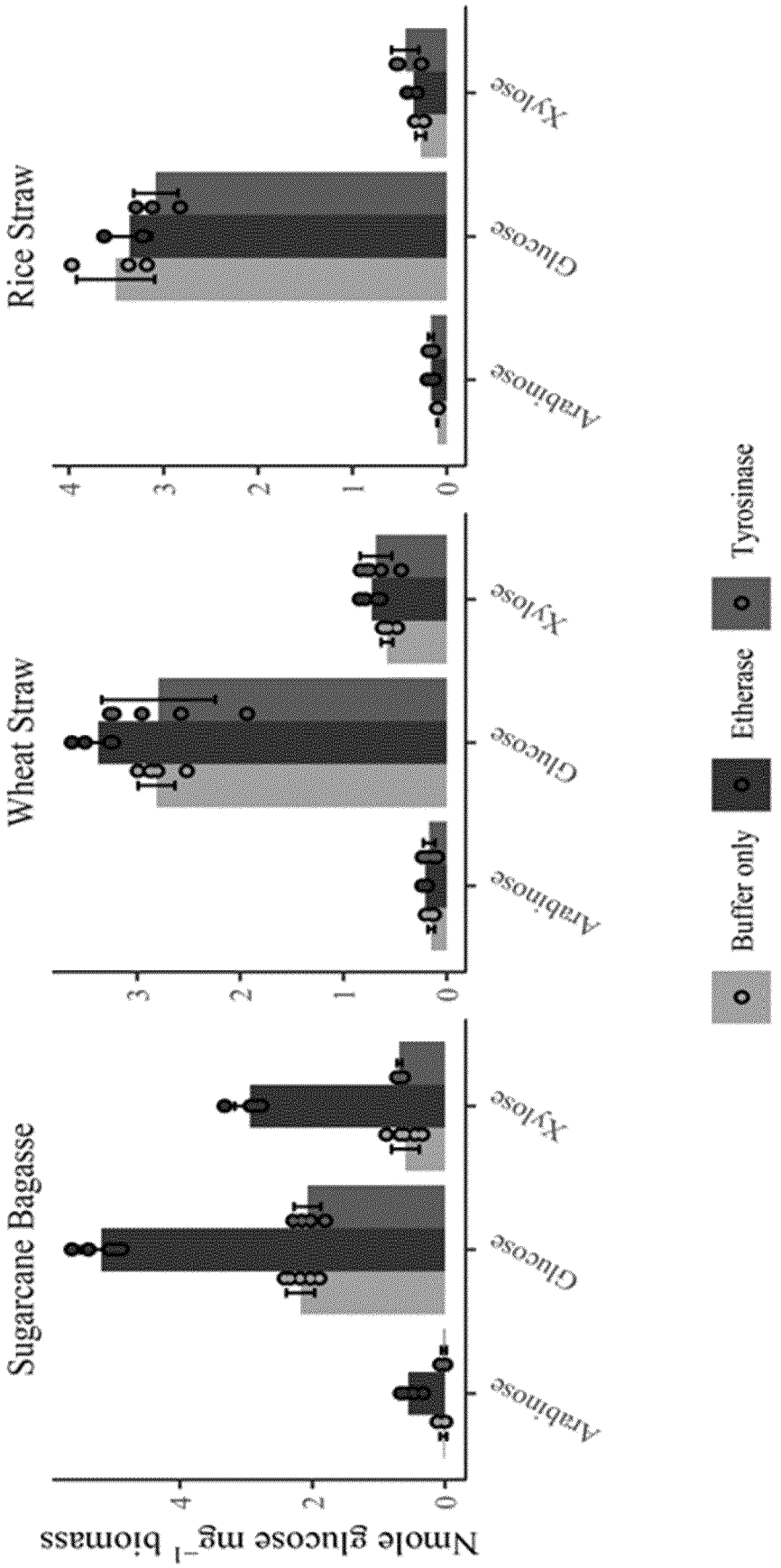

Nmole glucose mg⁻¹ biomass
Sugarcane Bagasse
Wheat Straw
Rice Straw
Arabinose
Glucose
Xylose
Buffer only
Etherase
Tyrosinase

FIG. 6A

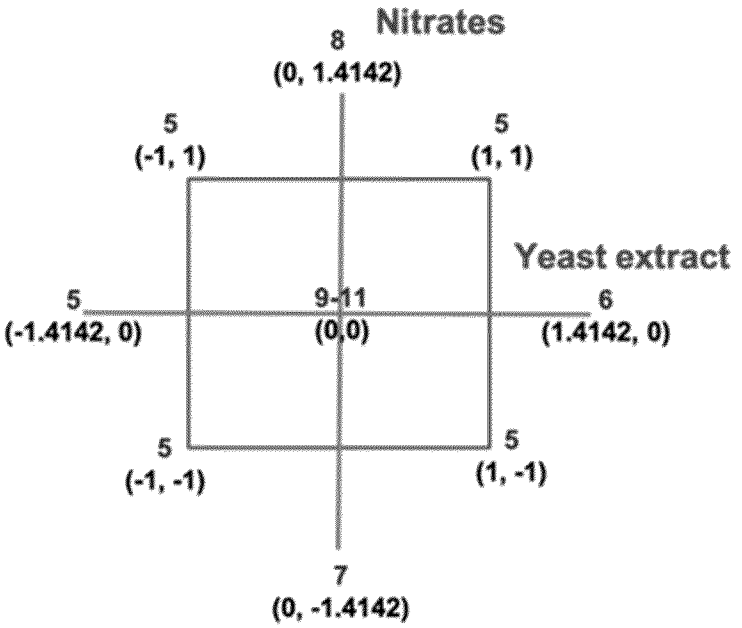

Nitrates
8
(0, 1.4142)
5
(-1, 1)
5
(1, 1)
Yeast extract
5
(-1.4142, 0)
9-11
(0,0)
6
(1.4142, 0)
5
(-1, -1)
5
(1, -1)
7
(0, -1.4142)

FIG. 6B

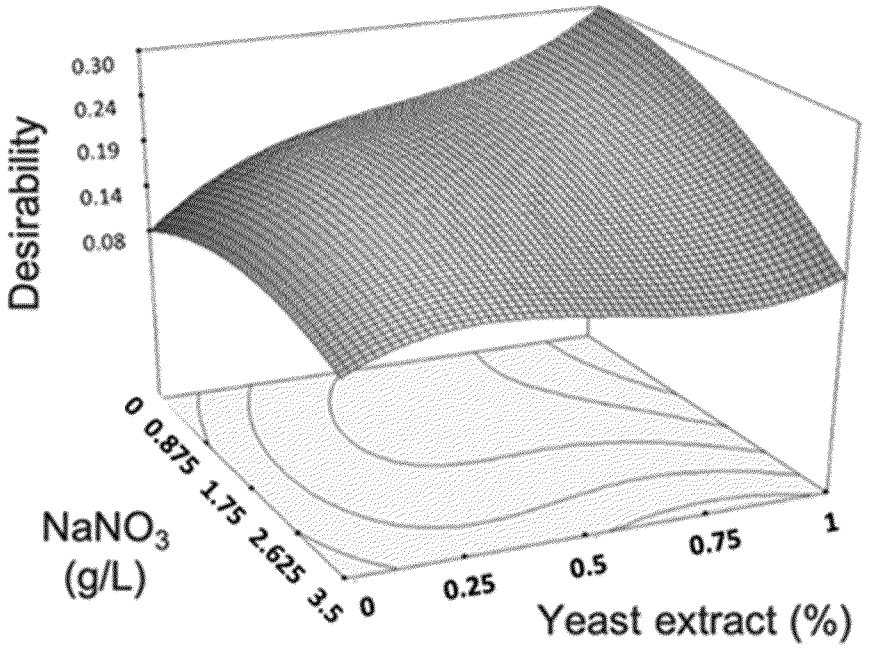

Desirability
0.30
0.24
0.19
0.14
0.08
NaNO3
(g/L)
0
0.875
1.75
2.625
3.5
Yeast extract (%)
0
0.25
0.5
0.75
1

FIG. 7A

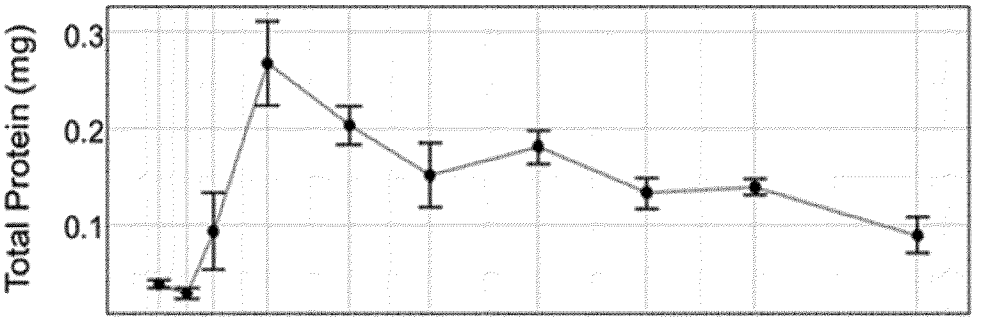


FIG. 7B

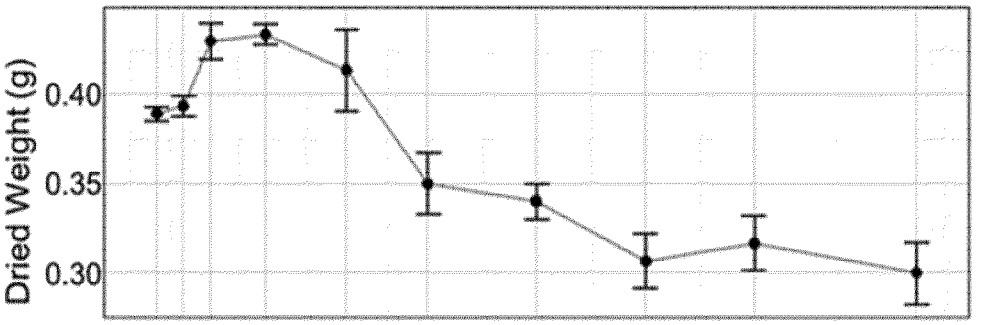


FIG. 7C

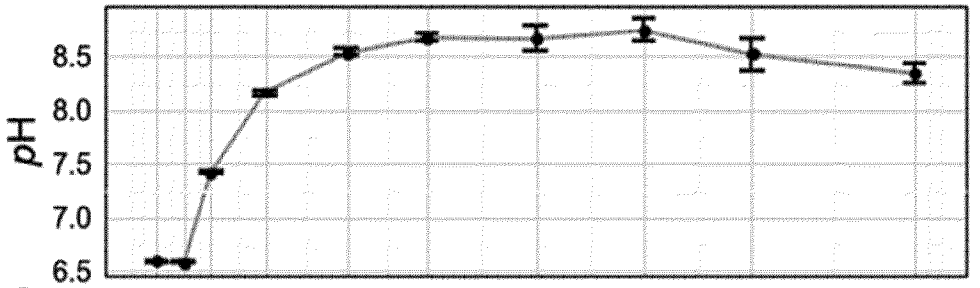


FIG. 7D

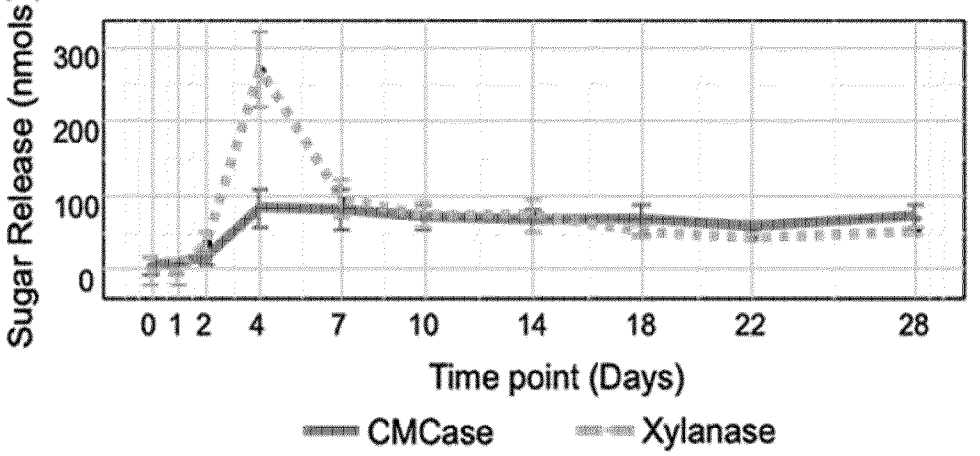


FIG. 8A

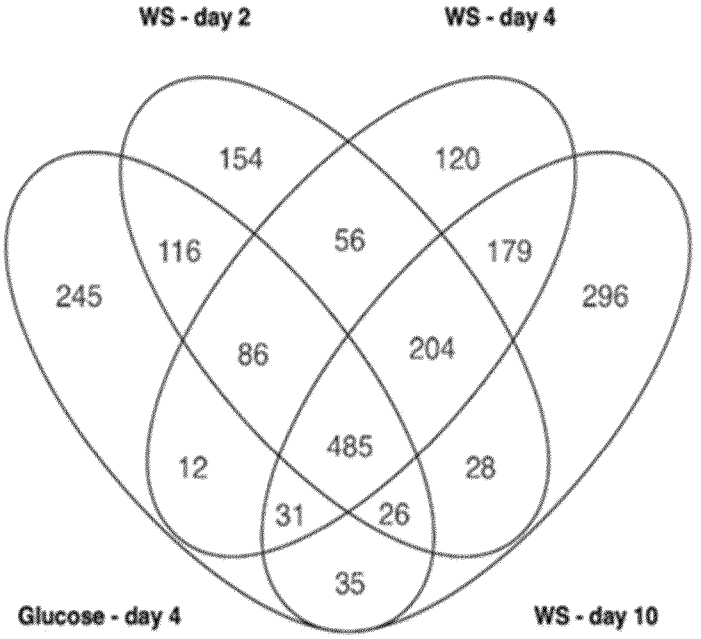

WS - day 2
WS - day 4
154
120
116
56
179
245
296
86
204
485
12
28
31
26
35
Glucose - day 4
WS - day 10

FIG. 8B

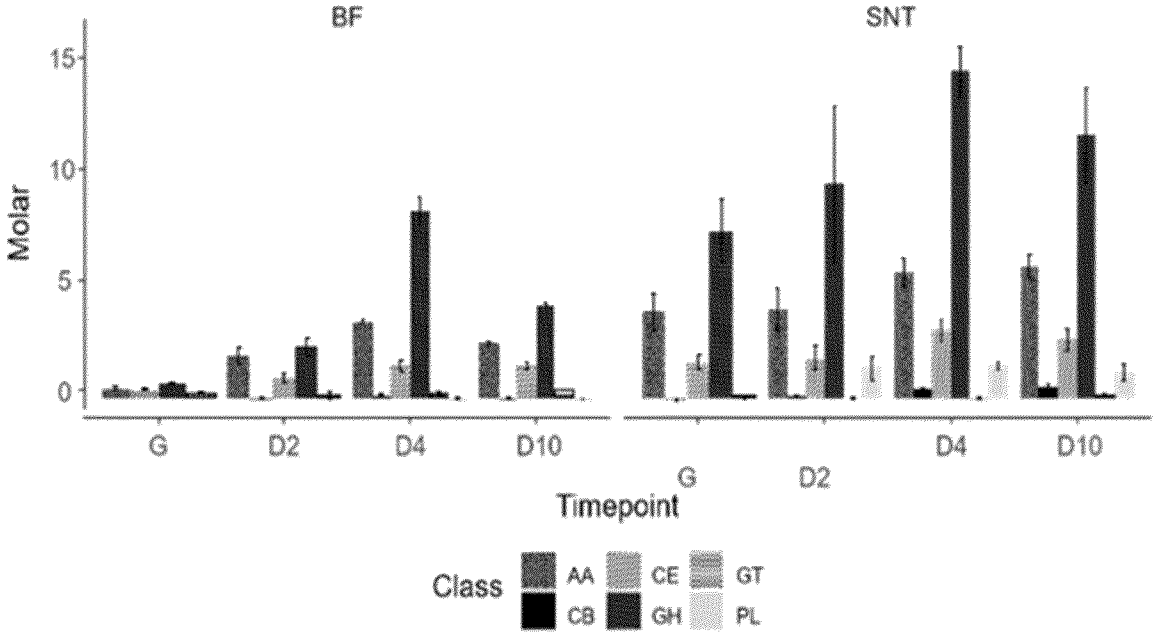

BF
SNT
Molar
15
10
5
0
G
D2
D4
D10
Timepoint
Class
AA
CB
CE
GH
GT
PL

FIG. 9

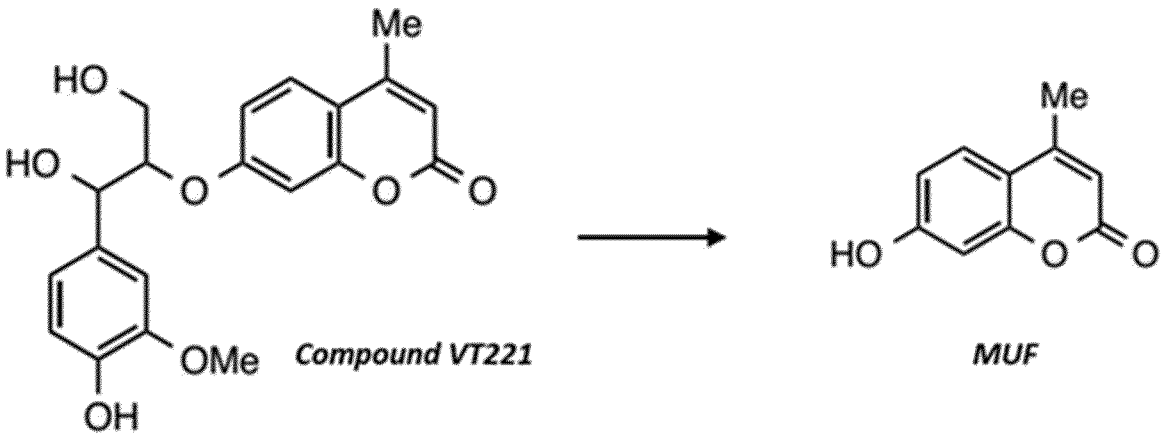


FIG. 10A

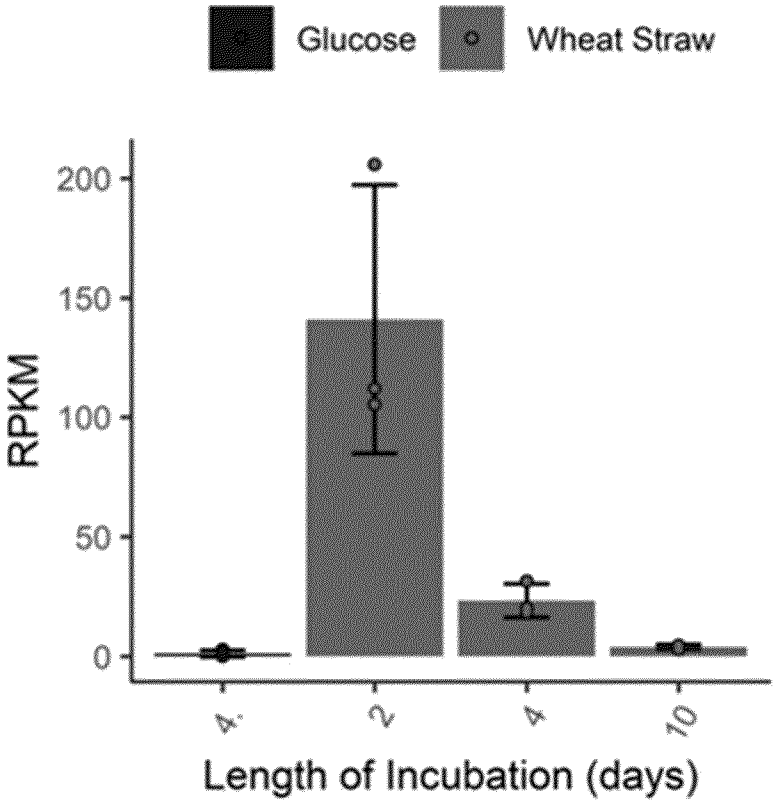


FIG. 10B

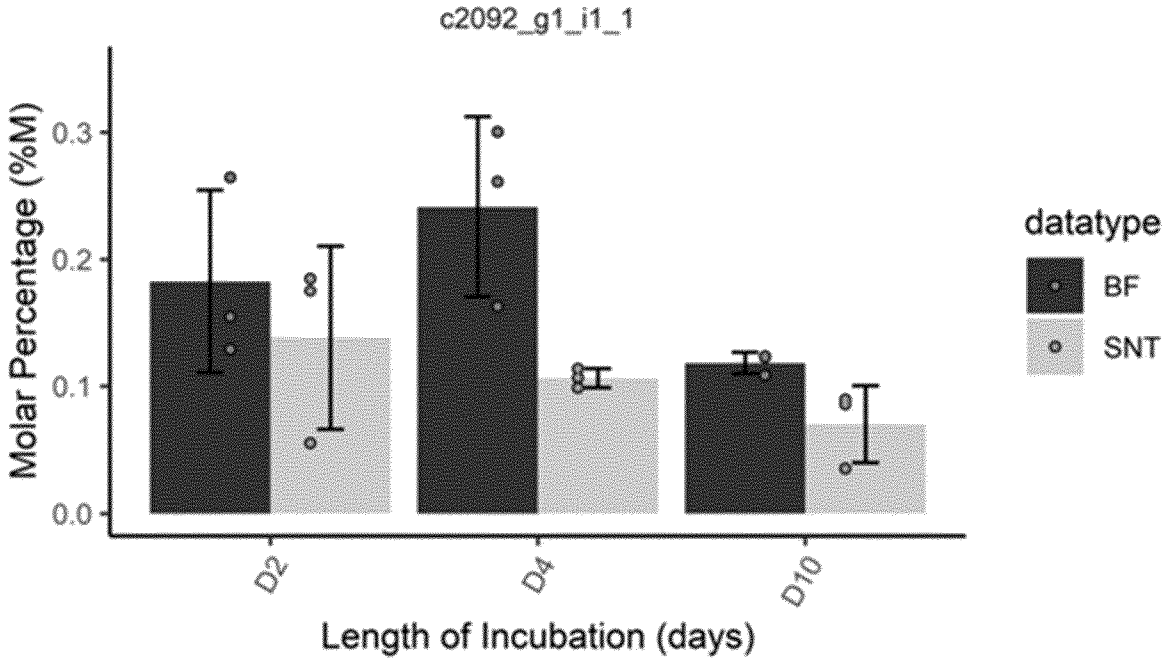


FIG. 11

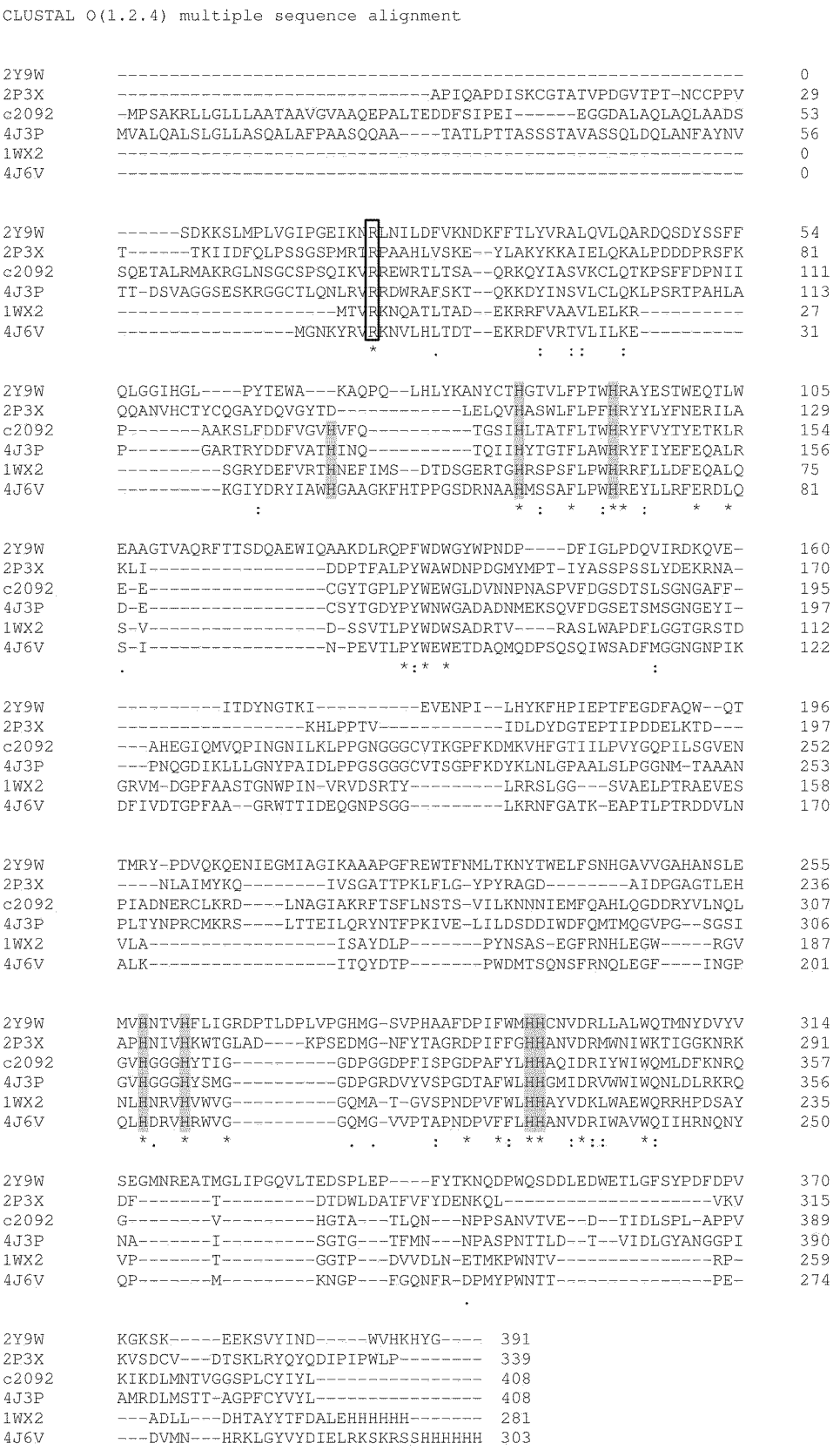

CLUSTAL O(1.2.4) multiple sequence alignment
2Y9W
2P3X
c2092
4J3P
1WX2
4J6V

FIG. 12

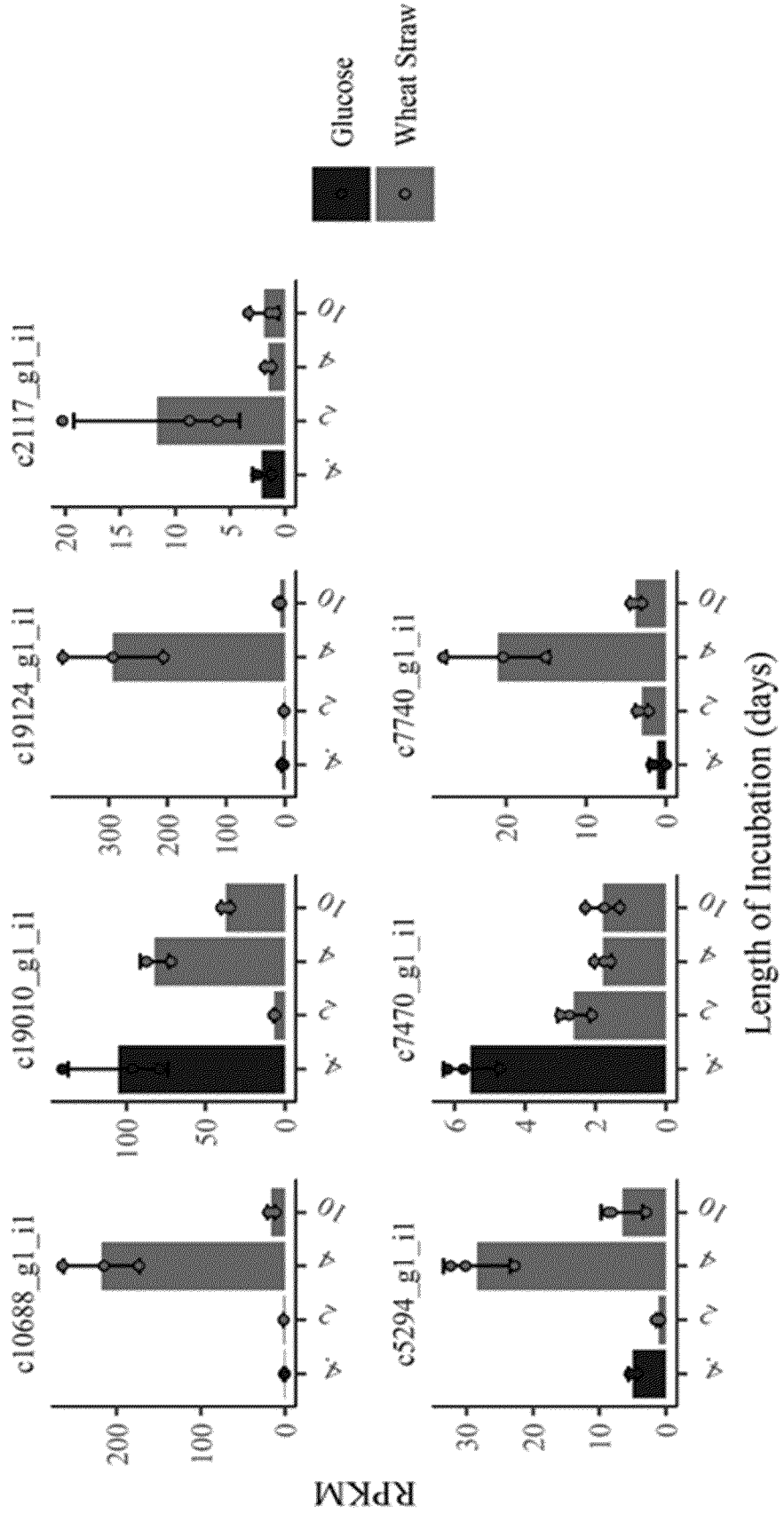

Glucose
Wheat Straw
c10688_g1_i1
c19010_g1_i1
c19124_g1_i1
c2117_g1_i1
c5294_g1_i1
c7470_g1_i1
c7740_g1_i1
RPKM
Length of Incubation (days)
4
2
4
10

FIG. 13A

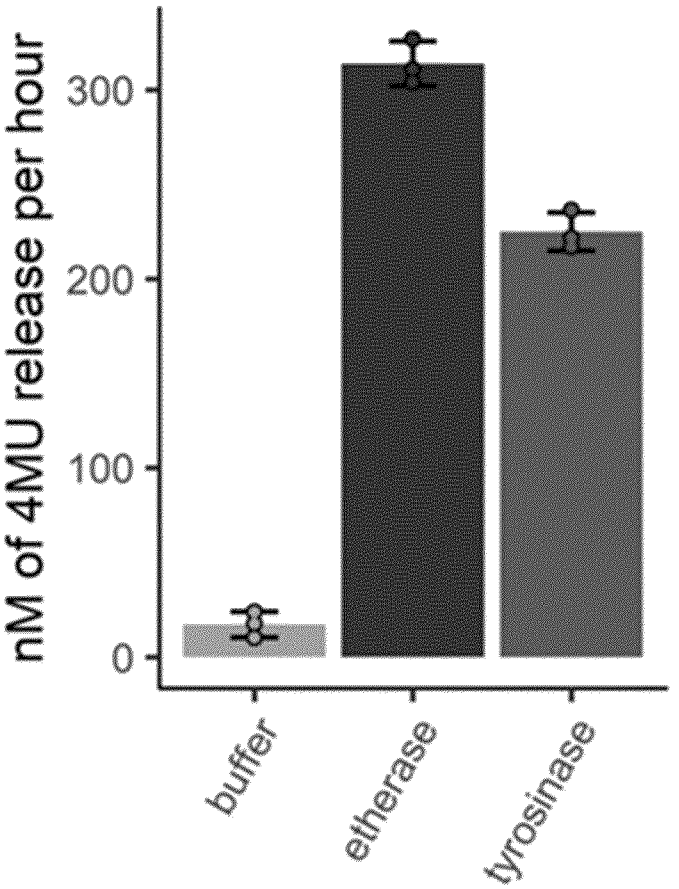
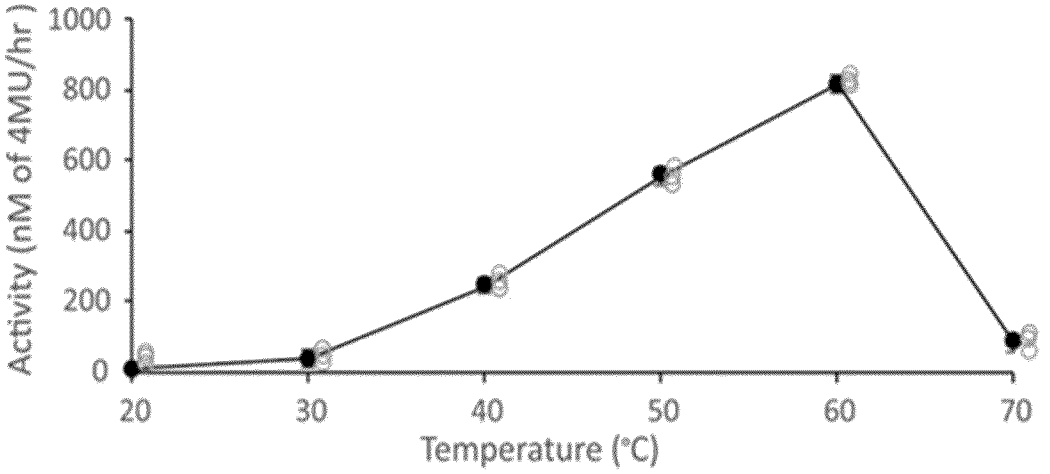

nM of 4MU release per hour
300
200
100
0
buffer
etherase
tyrosinase

FIG. 13B



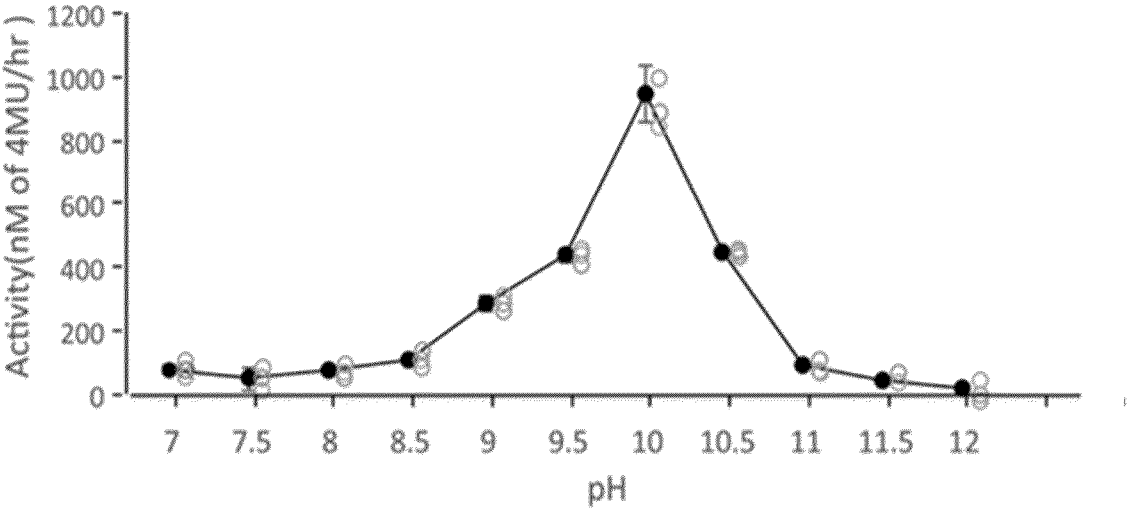


FIG. 13C

FIG. 14A

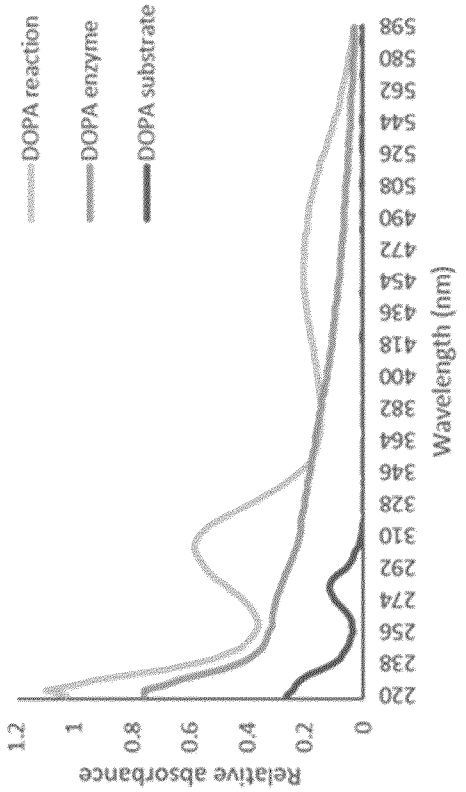


FIG. 14B

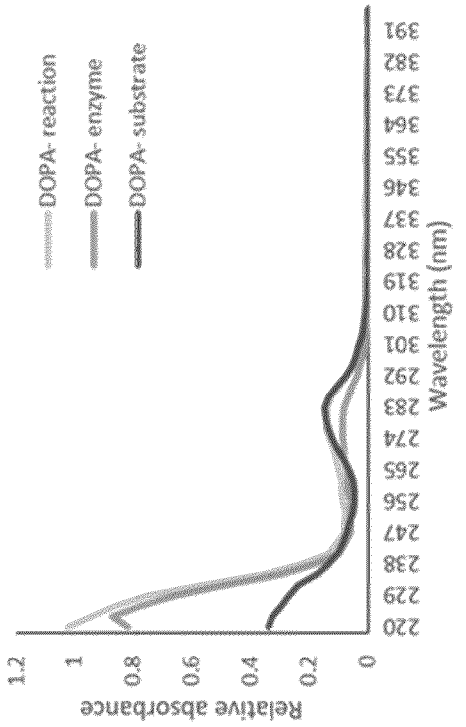


FIG. 14C

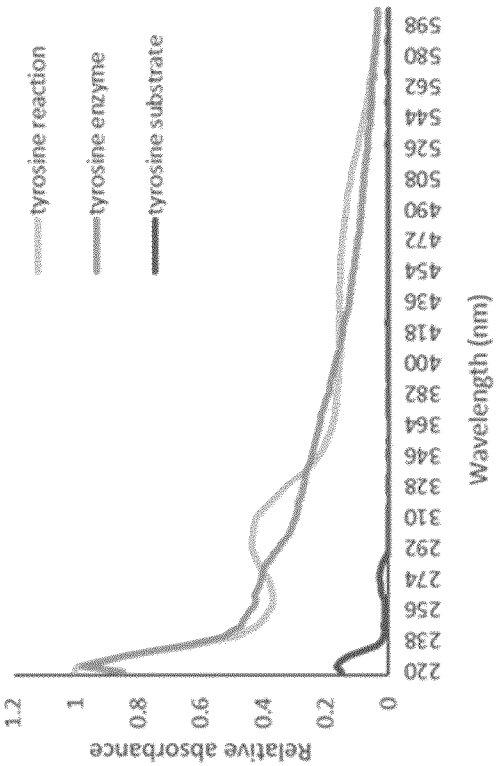


FIG. 14D

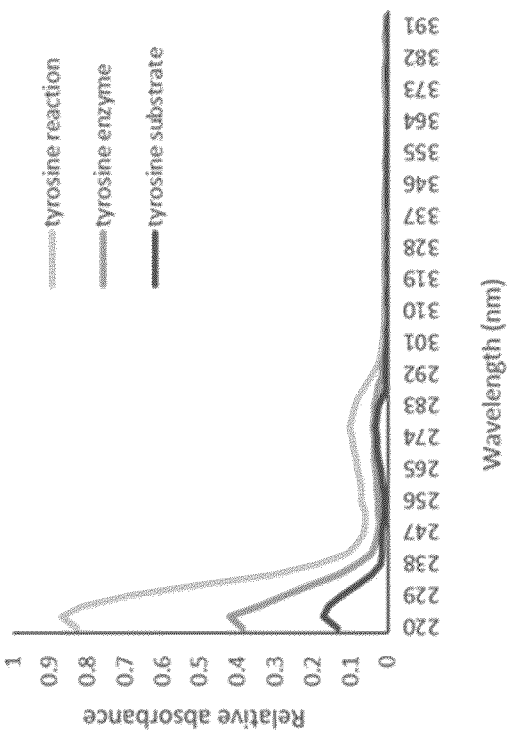


FIG. 15A

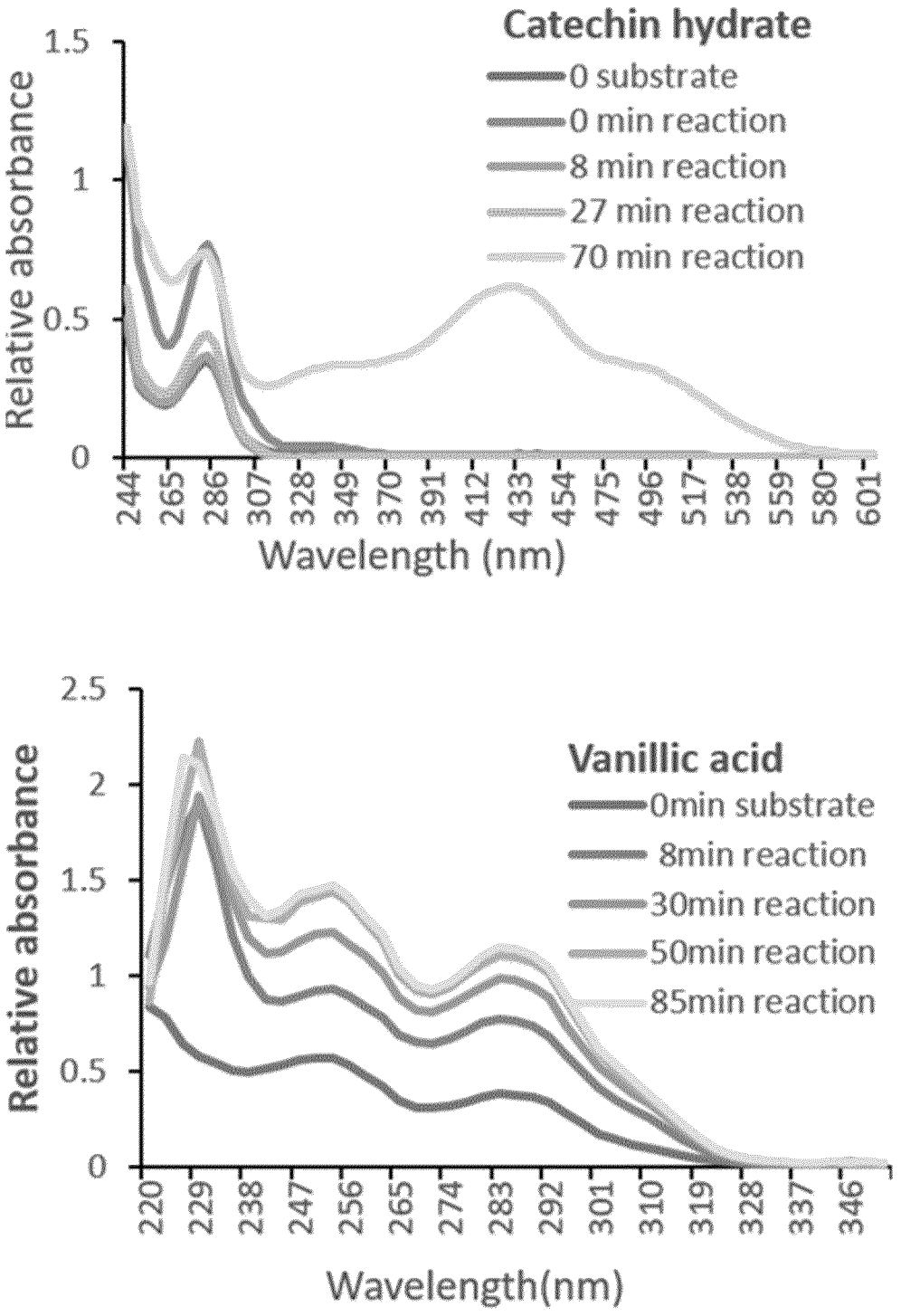

Catechin hydrate
0 substrate
0 min reaction
8 min reaction
27 min reaction
70 min reaction
Relative absorbance
1.5
1
0.5
0
244
265
286
307
328
349
370
391
412
433
454
475
496
517
538
559
580
601
Wavelength (nm)
Vanillic acid
0min substrate
8min reaction
30min reaction
50min reaction
85min reaction
Relative absorbance
2.5
2
1.5
1
0.5
0
220
229
238
247
256
265
274
283
292
301
310
319
328
337
346
Wavelength(nm)

FIG. 15B

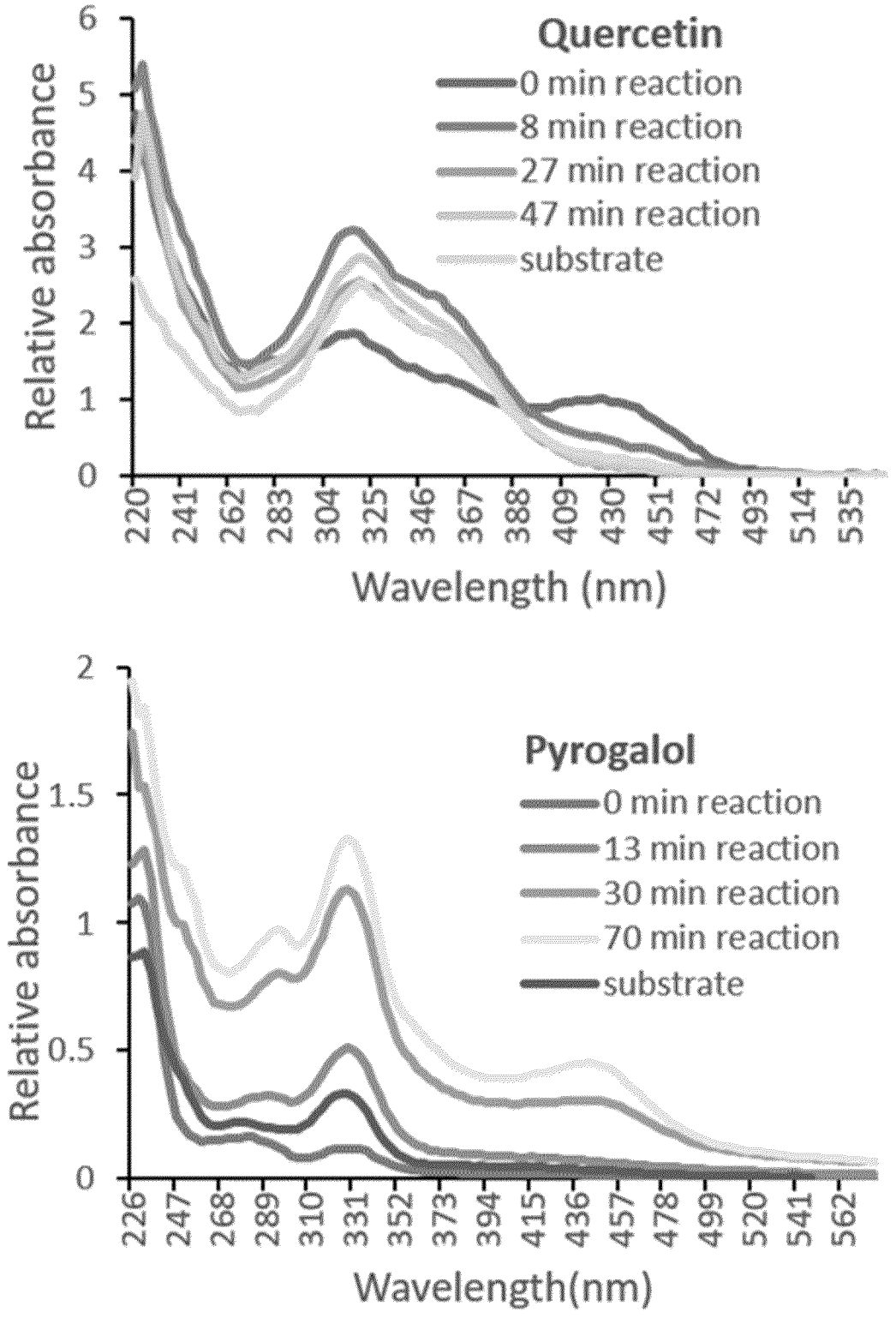

Quercetin
0 min reaction
8 min reaction
27 min reaction
47 min reaction
substrate
Relative absorbance
Wavelength (nm)
Pyrogalol
0 min reaction
13 min reaction
30 min reaction
70 min reaction
substrate
Relative absorbance
Wavelength(nm)

FIG. 15C

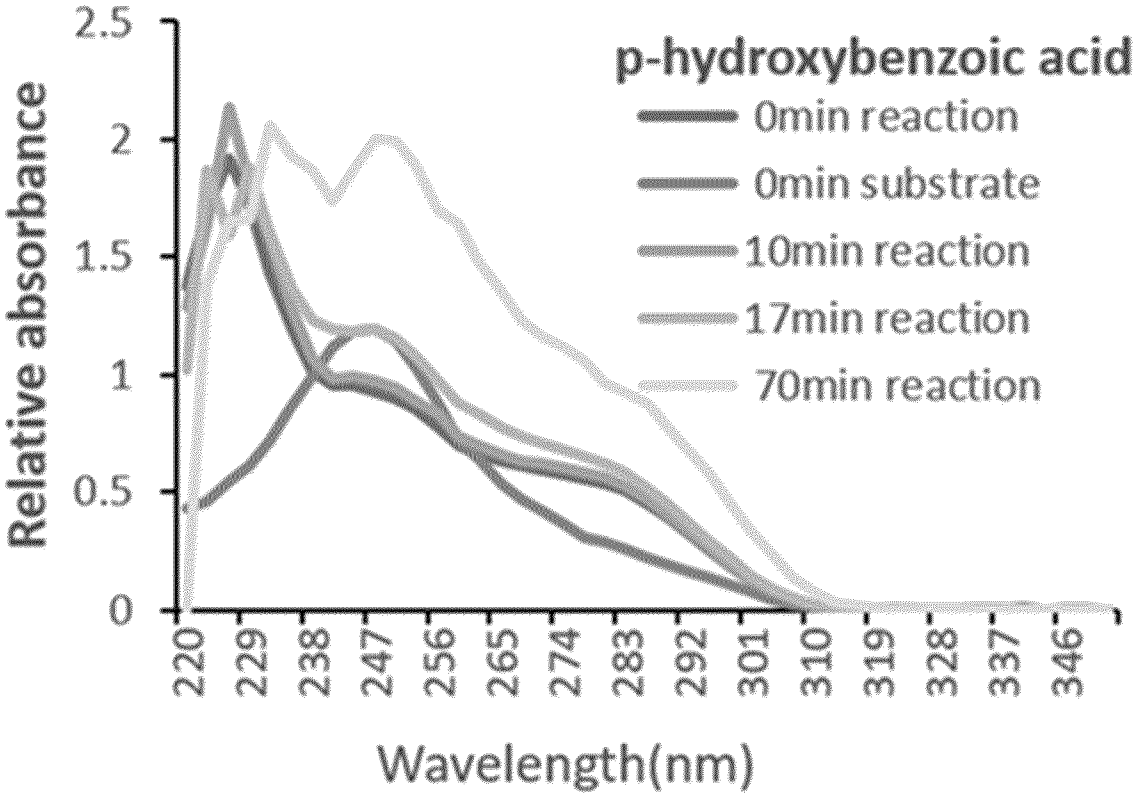

p-hydroxybenzoic acid
0min reaction
0min substrate
10min reaction
17min reaction
70min reaction
Relative absorbance
2.5
2
1.5
1
0.5
0
220
229
238
247
256
265
274
283
292
301
310
319
328
337
346
Wavelength(nm)

FIG. 16

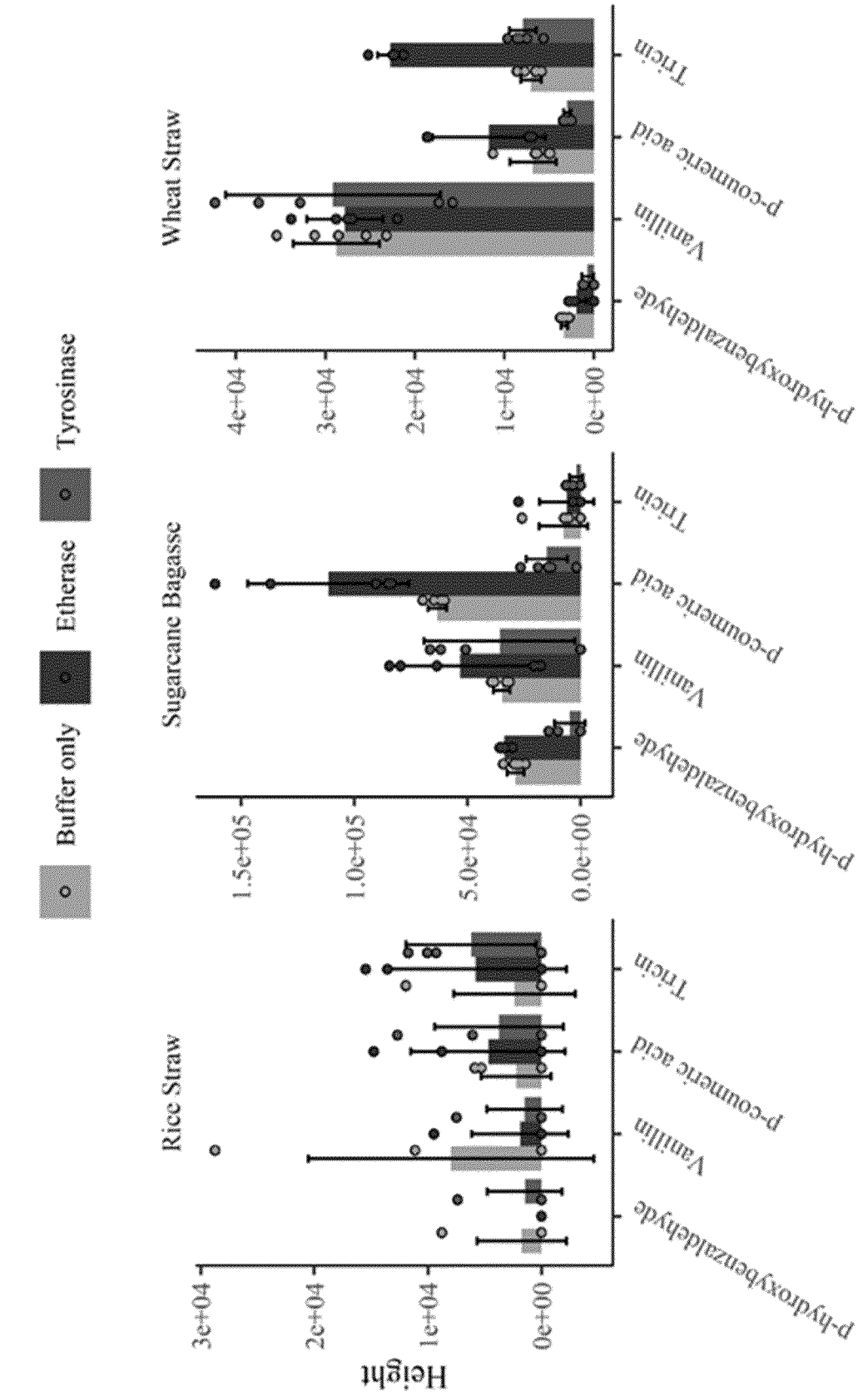

Buffer only
Etherase
Tyrosinase
Rice Straw
Sugarcane Bagasse
Wheat Straw
Height
0e+00
1e+04
2e+04
3e+04
0.0e+00
5.0e+04
1.0e+05
1.5e+05
0c+00
1c+04
2e+04
3e+04
4e+04
p-hydroxybenzaldehyde
Vanillin
p-coumeric acid
Tricin

FIG. 17A

FIG. 17B

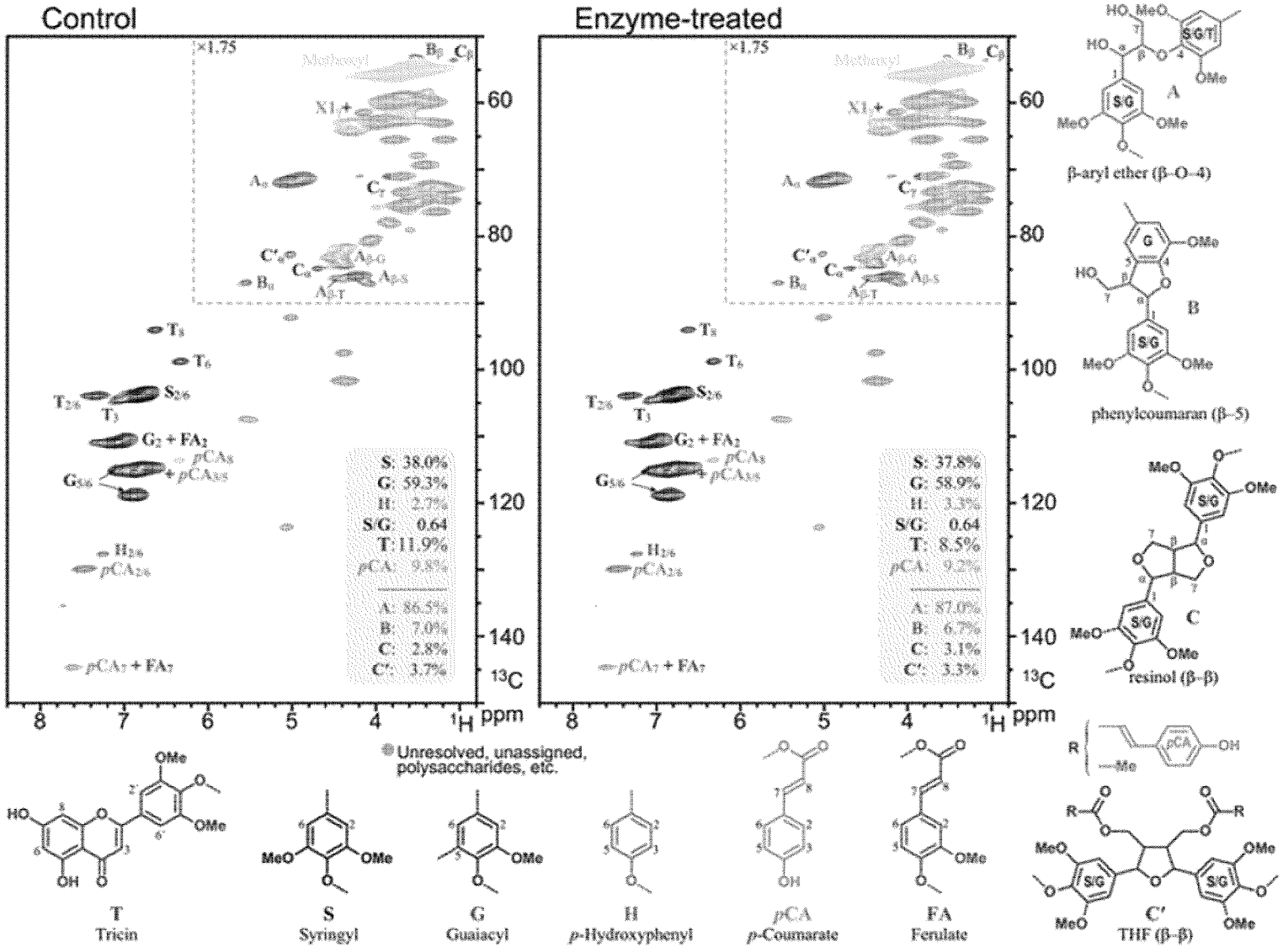
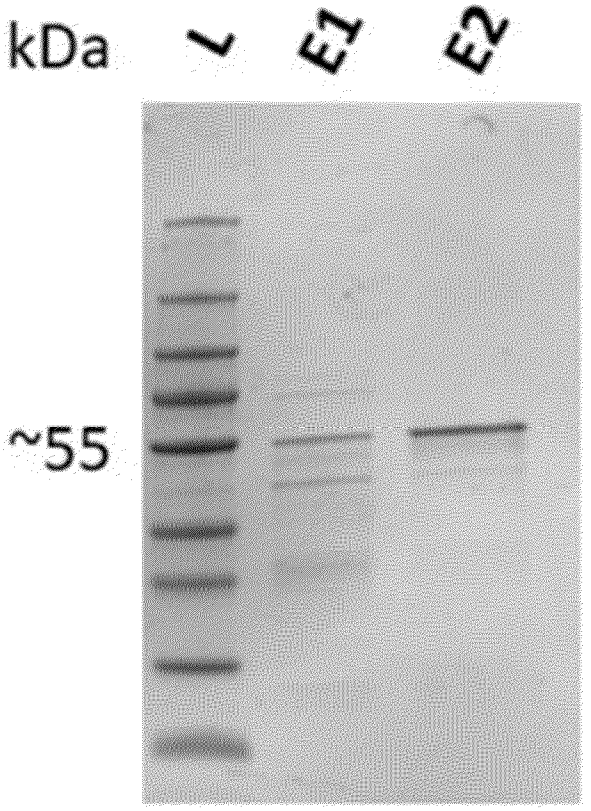


FIG. 18



BETA-ETHERASES FOR LIGNIN DEPOLYMERISATION

FIELD OF THE DISCLOSURE

[0001] The present application relates to nucleic acids encoding polypeptides with β-etherase activity; polypeptides with β-etherase activity; vectors comprising said nucleic acids for the production of recombinant β-etherase; cells, for example microbial cells, transformed with nucleic acids encoding β-etherase activity and vectors including nucleic acids encoding β-etherases; a composition comprising β-etherases suitable for processing lignocellulose; and a method that uses β-etherases or compositions comprising β-etherases in the processing of lignocellulose and related polysaccharides.

GOVERNMENT RIGHTS

[0002] This invention was made with government support under DE-SC0018409 awarded by the US Department of Energy. The government has certain rights in the invention.

BACKGROUND TO THE DISCLOSURE

[0003] The plant cell wall is composed of cellulose, hemicelluloses, pectic polysaccharides, and lignin, and is collectively termed lignocellulose. Photosynthetically fixed carbon in lignocellulose is produced in vast quantities on the Earth's surface. Its conversion into liquid transportation fuel represents a potential source of renewable energy with diverse feedstocks, including agricultural residues, municipal waste, and dedicated low-input crops. Effective utilization of lignocellulose, nevertheless, remains a challenge, as the extraction of fermentable sugars for biofuel production requires intensive physico-chemical pretreatments and high loadings of enzyme cocktails. A key factor of this recalcitrance to degradation is the presence of lignin, a heterogeneous, hydrophobic aromatic polymer that encases the cellulose and hemicellulose, blocking enzyme accessibility and impeding cellulase activity.

[0004] Lignin is synthesised by plants through the oxidative coupling of three hydroxycinnamyl alcohols: coniferyl alcohol, sinapyl alcohol and p-coumaryl alcohol, generating β—O—4, 4—O—5, β-5, β-1, 5-5 and β-β inter-unit linkages in β-ether, biphenyl ether, phenylcoumaran, spirodienone, biphenyl, and resinol units, respectively. Lignin requires a high redox potential to be oxidatively attacked. Recalcitrance to degradation is further enhanced as lignin has no defined repeat structure. The β—O—4 (or β-aryl) ether linkage is the most abundant linkage in the lignin macromolecule; its cleavage results in substantial lignin depolymerization.

[0005] Enzymes for depolymerising lignin are known and disclosed in US2019/048329 and include dehydrogenases, glutathione lyases and β-etherases which attack β—O—4 ether linkages. The β-etherase activity disclosed in US2019/048329 requires the co-substrates NAD^+ and glutathione.

[0006] Tricin, [5,7-dihydroxy-2-(4-hydroxy-3,5-dimethoxyphenyl)-4H-chromen-4-one], an O-methylated flavone, forms part of the structure of lignin from monocot plants including wheat, rice, sugar cane, and palms. Tricin has only been observed incorporated into the lignin structure via 4—O—β linkages, having arisen from the radical coupling of the flavone at its 4'—O—position with the monolignol at its β-position.

[0007] Tricin is recognized as a valuable human health compound due to its antioxidant, anti-aging, anti-cancer, and cardio-protective potential. Tricin may be present as its parent compound that may be released by solvent extraction from a variety of monocotyledons such as wheat (Triticum aestivum), oat bran (Avena sativa), bamboo (Leleba oldhami), sugarcane (Saccharum officinarum), and maize (Zea mays), and has been observed in quantities of up to 3.3% wt of lignin from wheat straw.

[0008] This disclosure characterises a copper-containing β-etherase that can cleave the β-aryl ether linkage of lignin and which is secreted from the fungus Parascedosporium when growing on wheat straw. The disclosed β-etherase has no requirement for NAD^+ and/or glutathione and was found to readily cleave tricin from wheat straw, also enhancing the saccharification of lignocellulosic biomass when used in combination with cellulolytic enzymes.

STATEMENTS OF THE INVENTION

[0009] According to an aspect of the invention there is provided an isolated nucleic acid molecule encoding a β-etherase polypeptide wherein said polypeptide comprises copper and further wherein the activity of said polypeptide is independent of NAD^+ and/or glutathione.

[0010] Lignin, the major component of lignocellulosic plant biomass, is an organic heterologous polymer comprising covalently linked phenylpropanoid units and consist essentially of crosslinked methoxylated derivatives of benzene such as p-coumaryl, coniferyl, and sinapyl alcohols. Exemplary phenylpropanoid units derived from the alcohols are p-hydroxyphenyl, guaiacyl, and syringyl units respectively. The phenylpropanoid units can be linked to other phenylpropanoid units through bonds such as β—O—4, 4—O—5, β-5, β-1, 5-5 and β-β inter-unit linkages. β—O—4 ether bonds account for 45-60% of linkages present in lignin. Flavonoid units such as tricin can be incorporated into lignin via 4—O—β ether bonds.

[0011] β-etherase activity in the context of this application refers to the capability to cleave β-aryl ether (β—O—4) bonds in lignin that link one phenylpropanoid unit to another phenylpropanoid unit or to flavonoid units such as tricin.

[0012] In order to optimize expression levels in recombinant host cells, codon optimisation of the nucleic acid sequence to be expressed may be required to convert a natural sequence to a non-natural sequence that encodes substantially the same polypeptide and would be optimally expressed in a heterologous host cell. Codon optimisation is known in the art and increases translational efficiency in the desired host organism and replace codons of low frequency with codons of high frequency.

[0013] In a preferred embodiment of the invention, the said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0014] i) a nucleotide sequence as set forth in SEQ ID NO: 1;

[0015] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0016] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which

hybridizes under stringent hybridisation conditions to sequence set forth in SEQ ID NO 1;

[0017] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 9;

[0018] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0019] Hybridization of a nucleic acid molecule occurs when two complementary nucleic acid molecules undergo an amount of hydrogen bonding to each other. The stringency of hybridization can vary according to the environmental conditions surrounding the nucleic acids, the nature of the hybridization method, and the composition and length of the nucleic acid molecules used. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed in Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); and Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes Part I, Chapter 2 (Elsevier, New York, 1993). The T_m is the temperature at which 50% of a given strand of a nucleic acid molecule is hybridized to its complementary strand. The following is an exemplary set of hybridization conditions and is not limiting:

[0020] Very High Stringency (allows sequences that share at least 90% or 95% identity to hybridize)

[0021] Hybridization: 5x SSC at 65° C. for 16 hours

[0022] Wash twice: 2x SSC at room temperature (RT) for 15 minutes each

[0023] Wash twice: 0.5x SSC at 65° C. for 20 minutes each

[0024] High Stringency (allows sequences that share at least 80% identity to hybridize)

[0025] Hybridization: 5x-6x SSC at 65-70° C. for 16-20 hours

[0026] Wash twice: 2x SSC at RT for 5-20 minutes each

[0027] Wash twice: 1x SSC at 55-70° C. for 30 minutes each

[0028] Low Stringency (allows sequences that share at least 50% identity to hybridize)

[0029] Hybridization: 6x SSC at RT to 55° C. for 16-20 hours

[0030] Wash at least twice: 2x-3x SSC at RT to 55° C. for 20-30 minutes each.

[0031] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0032] i) a nucleotide sequence as set forth in SEQ ID NO: 2;

[0033] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0034] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 2;

[0035] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 10;

[0036] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0037] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0038] i) a nucleotide sequence set forth in SEQ ID NO: 3;

[0039] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0040] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 3;

[0041] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO 11;

[0042] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0043] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0044] i) a nucleotide sequence as set forth in SEQ ID NO 4;

[0045] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0046] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 4;

[0047] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO 12:

[0048] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0049] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0050] i) a nucleotide sequence as set forth in SEQ ID NO 5:

[0051] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0052] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 5;

[0053] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence as represented in SEQ ID NO 13;

[0054] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion

or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0055] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0056] i) a nucleic acid sequences as set forth in SEQ ID NO 6;

[0057] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0058] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 6;

[0059] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence as set forth in SEQ ID NO 14;

[0060] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0061] In a preferred embodiment of the invention said isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

[0062] i) a nucleotide sequence as set forth in SEQ ID NO: 7;

[0063] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0064] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions set forth in SEQ ID NO 7;

[0065] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence as set forth SEQ ID NO 15;

[0066] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0067] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0068] i) a nucleotide sequence as set forth in SEQ ID NO 8;

[0069] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0070] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 8;

[0071] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 16;

[0072] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0073] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0074] i) a nucleotide sequence as set forth in SEQ ID NO 18 or 17;

[0075] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0076] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 18 or 17;

[0077] iv) a nucleotide sequence that encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO 26;

[0078] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0079] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0080] i) a nucleotide sequence as set forth in SEQ ID NO 19;

[0081] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0082] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 19;

[0083] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 27;

[0084] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0085] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0086] i) a nucleotide sequence as set forth in SEQ ID NO 20;

[0087] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0088] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 20;

[0089] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 28;

[0090] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0091] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0092] i) a nucleotide sequence as set forth in SEQ ID NO 21;

[0093] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0094] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 21;

[0095] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 29;

[0096] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0097] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0098] i) a nucleotide sequence as set forth in SEQ ID NO 22;

[0099] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0100] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 22;

[0101] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 30;

[0102] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0103] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0104] i) a nucleotide sequence as set forth in SEQ ID NO 23;

[0105] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0106] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 23;

[0107] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 31;

[0108] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0109] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0110] i) a nucleotide sequence as set forth in SEQ ID NO 24;

[0111] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0112] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 24;

[0113] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 32;

[0114] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0115] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0116] i) a nucleotide sequence as set forth in SEQ ID NO 24;

[0117] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0118] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 24;

[0119] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 32;

[0120] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0121] In a preferred embodiment of the invention said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

[0122] i) a nucleotide sequence as set forth in SEQ ID NO 25;

[0123] ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

[0124] iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to the sequence set forth in SEQ ID NO 25;

[0125] iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO 33;

[0126] v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

[0127] The presence of a peptide signal sequence encoded by part of the nucleic acid sequence set forth in SEQ ID NO 1-8 which is located at the N-terminus of the amino acid sequences set forth in SEQ ID NO 9-16, may result in inefficient expression of the protein in an alternative expression host cell. Therefore, typically, the endogenous host specific signal sequence is either replaced with the expression host specific peptide signal sequence or with an ATG codon. The nucleotide sequences set forth in sequence IDs 17-25 represent the nucleotide sequence lacking the signal sequence or an ATG start codon at the 5'-end of the nucleotide sequence and correspondingly, the amino acid sequences set forth in SEQ IDs No 26-33 are lacking the N-terminal signal sequence or a methionine as the first amino acid at the N-terminus of the amino acid sequence. Thus, nucleotide sequences set forth in SEQ ID NO 17-25 comprising an

ATG as the first codon at the 5'-end or amino acid sequences set forth in SEQ ID NO 26-33 comprising a methionine as the first amino acid of the N-terminus are also claimed.

[0128] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 1 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0129] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 2 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0130] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 3 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0131] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 4 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity

[0132] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 5 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0133] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 6 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0134] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 7 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0135] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 8 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0136] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 17 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0137] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 18 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0138] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 19 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0139] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 20 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0140] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 21 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0141] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 22 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0142] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 23 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0143] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 24 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0144] In a preferred embodiment of the invention said nucleic acid molecule comprises or consists of a nucleotide sequence as set forth in SEQ ID NO: 25 wherein said nucleic acid molecule encodes a polypeptide with β-etherase activity.

[0145] According to a further aspect of the invention there is provided an isolated β-etherase polypeptide wherein said polypeptide comprises copper and further wherein the activity of said polypeptide is independent of NAD^+ and/or glutathione.

[0146] In a preferred embodiment of the invention said β-etherase polypeptide comprises two copper binding sites comprising the motif:

[0147] Copper binding site No 1: H—X(1-7)—H—X(1-8)—H and site No 2:H— X(1-3)—H—X(22-25)—H;

[0148] wherein X is any amino acid and H is histidine. The numerical range X (1-7), X (1-8), X (1-3) and X (22-25) denotes the number of amino acid residues between the histidines e.g., H—X (1-3)—H contains three amino acid residues between the two histidines. Variations to this motif are shown in FIG. **11**.

[0149] In a preferred embodiment of the invention said polypeptide has β-etherase activity in the absence of NAD^+ and glutathione.

[0150] In a further preferred embodiment of the invention said isolated β-etherase polypeptides share at least 23% sequence identity over the full-length sequence set forth in SEQ ID NO 9 or 26

[0151] In a further preferred embodiment of the invention said isolated β-etherase polypeptides share between 23-45% sequence identity over the full-length sequence set forth in SEQ ID NO 9 or 26.

[0152] In a further preferred embodiment of the invention said isolated β-etherase polypeptides share at least 23%, 24%, 25%, 30%, 35%, 37%, 38%, 39%, 40%, 41%, 44% and 45% sequence identity over the full-length sequence set forth in SEQ ID NO 9 or 26.

[0153] In an alternative further preferred embodiment of the invention said isolated β-etherase polypeptides share at least 50% sequence identity over the full-length sequence set forth in SEQ ID NO 9 or 26.

[0154] In an alternative further preferred embodiment of the invention said isolated β-etherase polypeptides share

between 50-88% sequence identity over the full-length sequence set forth in SEQ ID NO 9 or 26.

[0155] In an alternative further preferred embodiment of the invention said isolated β-etherase polypeptides share at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% and 99% sequence identity over the full-length sequence set forth in SEQ ID NO 9 or 26.

[0156] In a preferred embodiment of the invention said isolated polypeptide is selected from the group consisting of:

[0157] i) a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 9 or 26;

[0158] ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 9 or 26 and which has β-etherase activity.

[0159] In a preferred embodiment of the invention said isolated polypeptide is selected from the group consisting of:

[0160] i) a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 10 or 27;

[0161] ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 10 or 27 and which has β-etherase activity.

[0162] According to an aspect of the invention there is provided an isolated polypeptide selected from the group consisting of:

[0163] i) a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 11 or 28;

[0164] ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 11 or 28 and which has β-etherase activity.

[0165] In a preferred embodiment of the invention said isolated polypeptide is selected from the group consisting of:

[0166] i) a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 12 or 29;

[0167] ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 12 or 29 and which has β-etherase activity.

[0168] In a preferred embodiment of the invention said isolated polypeptide is selected from the group consisting of:

[0169] i) a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 13 or 30;

[0170] ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 13 or 30 and which has β-etherase activity.

[0171] In a preferred embodiment of the invention said isolated polypeptide is selected from the group consisting of:

[0172] i) a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 14 or 31;

[0173] ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 14 or 31 and which has β-etherase activity.

[0174] In a preferred embodiment of the invention said isolated polypeptide is selected from the group consisting of:

[0175] i) a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 15 or 32;

[0176] ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 15 or 32 and which has β-etherase activity.

[0177] In a preferred embodiment of the invention said isolated polypeptide is selected from the group consisting of:

[0178] i) a polypeptide comprising or consisting of an amino acid sequence set forth in SEQ ID NO: 16 or 33

[0179] ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 16 or 33 and which has β-etherase activity.

[0180] A modified polypeptide as herein disclosed may differ in amino acid sequence by one or more substitutions, additions, deletions, truncations that may be present in any combination. Among preferred variants are those that vary from a reference polypeptide by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid by another amino acid of like characteristics. The following non-limiting list of amino acids are considered conservative replacements (similar): a) alanine, serine, and threonine; b) glutamic acid and aspartic acid; c) asparagine and glutamine d) arginine and lysine; e) isoleucine, leucine, methionine and valine and f) phenylalanine, tyrosine and tryptophan. Most highly preferred are variants that retain the same biological function and activity as the reference polypeptide from which it varies.

[0181] In a preferred embodiment of the invention the modified polypeptides have at least 23%, 24%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% identity, and at least 99% identity with the full-length amino acid sequence illustrated herein.

[0182] In a preferred embodiment of the invention the modified polypeptides have at least 23% identity with the full-length amino acid sequence illustrated herein.

[0183] In a preferred embodiment of the invention the modified polypeptides have at least 88% identity with the full-length amino acid sequence illustrated herein.

[0184] According to a further aspect of the invention there is provided a vector comprising a nucleic acid molecule according to the invention.

[0185] In a preferred embodiment of the invention the vector is an expression vector adapted for expression in a microbial host cell as herein disclosed.

[0186] Preferably the nucleic acid molecule in the vector is under the control of, and operably linked to, an appropriate promoter or other regulatory elements for transcription in a host cell such as a microbial, (e.g., bacterial, yeast), or plant cell. The vector may be a bi-functional expression vector which functions in multiple hosts.

[0187] According to a further aspect of the invention there is provided a host cell transformed or transfected with a nucleic acid molecule or vector according to the invention. In a preferred embodiment of the invention said cell is a heterologous host cell wherein said heterologous host cell does not naturally express a nucleic acid molecule according to the invention or vector comprising a nucleic acid molecule according to the invention.

[0188] In a further preferred embodiment of the invention said cell transformed or transfected with a nucleic acid molecule or vector according to the invention is a recombinant cell.

[0189] In the context of this application a recombinant cell defines a host organism cell comprising DNA from a different species e.g. expression of a nucleotide sequence from Parascedosporium species in an Aspergillus spp cell. In a preferred embodiment of the invention said cell is a microbial cell.

[0190] In a preferred embodiment said cell is selected from the group consisting of bacterial cell, yeast cell, fungal cell, insect cell and plant cell.

[0191] In a preferred embodiment said cell is a bacterial cell.

[0192] In a preferred embodiment of the invention said bacterial cell is an Escherichia coli cell.

[0193] In a preferred embodiment said transgenic is a fungal or yeast cell.

[0194] In a further preferred embodiment of the invention said fungal cell is an Aspergillus sp. cell

[0195] In a further preferred embodiment of the invention said fungal cell is an Aspergillus niger cell.

[0196] In a further preferred embodiment of the invention said fungal cell is not a Parascedosporium sp cell.

[0197] In a preferred embodiment of the invention said yeast cell is selected from the group consisting of Saccharomyces cerevisae, Schizosaccharomyces pombe or Pichia pastoris.

[0198] If microbial cells are used as organisms and in the process according to the invention they are grown or cultured in the manner with which the skilled worker is familiar, depending on the host organism. As a rule, microorganisms are grown in a liquid medium comprising a carbon source, usually in the form of sugars, a nitrogen source, usually in the form of organic nitrogen sources such as yeast extract or salts such as ammonium sulphate, trace elements such as salts of iron, copper, manganese and magnesium and, if appropriate, vitamins, at temperatures of between 0° C. and 100° C., preferably between 10° C. and 60° C., while gassing in oxygen.

[0199] The pH of the liquid medium can either be kept constant and regulated during the culturing period, or not. The cultures can be grown batchwise, semi-batchwise or continuously. Nutrients can be provided at the beginning of the fermentation or fed in semi-continuously or continuously. To this end, the organisms can advantageously be disrupted beforehand. In this process, the pH value is advantageously kept between pH 4 and 12, preferably between pH 6 and 9, especially preferably between pH 7 and 8.

[0200] The culture medium to be used must suitably meet the requirements of the strains in question. Descriptions of culture media for various microorganisms can be found in the textbook "Manual of Methods for General Bacteriology" of the American Society for Bacteriology (Washington D.C., USA, 1981).

[0201] As described above, these media which can be employed in accordance with the invention usually comprise one or more carbon sources, nitrogen sources, inorganic salts, vitamins and/or trace elements.

[0202] Preferred carbon sources are sugars, such as mono-, di- or polysaccharides. Examples of carbon sources are glucose, fructose, mannose, galactose, ribose, sorbose, ribulose, lactose, maltose, sucrose, raffinose, starch or cellulose. Sugars can also be added to the media via complex compounds such as molasses or other by-products from sugar refining. The addition of mixtures of a variety of carbon sources may also be advantageous. Other possible carbon sources are oils and fats such as, for example, soya oil, sunflower oil, peanut oil and/or coconut fat, fatty acids such as, for example, palmitic acid, stearic acid and/or linoleic acid, alcohols and/or polyalcohols such as, for example, glycerol, methanol and/or ethanol, and/or organic acids such as, for example, acetic acid and/or lactic acid.

[0203] Nitrogen sources are usually organic or inorganic nitrogen compounds or materials comprising these compounds. Examples of nitrogen sources comprise ammonia in liquid or gaseous form or ammonium salts such as ammonium sulphate, ammonium chloride, ammonium phosphate, ammonium carbonate or ammonium nitrate, nitrates, urea, amino acids, or complex nitrogen sources such as cornsteep liquor, soya meal, soya protein, yeast extract, meat extract, and others. The nitrogen sources can be used individually or as a mixture.

[0204] Inorganic salt compounds which may be present in the media comprise the chloride, phosphorus and sulphate salts of calcium, magnesium, sodium, cobalt, molybdenum, potassium, manganese, zinc, copper, and iron.

[0205] Inorganic sulphur-containing compounds such as, for example, sulphates, sulphites, dithionites, tetrathionates, thiosulfates, sulphides, or else organic sulphur compounds such as mercaptans and thiols may be used as sources of sulphur for the production of sulphur-containing fine chemicals and pathway intermediates, in particular of methionine.

[0206] Phosphoric acid, potassium dihydrogenphosphate or dipotassium hydrogenphosphate or the corresponding sodium-containing salts may be used as sources of phosphorus.

[0207] Chelating agents may be added to the medium in order to keep the metal ions in solution. Particularly suitable chelating agents comprise dihydroxyphenols such as catechol or protocatechuate and organic acids such as citric acid.

[0208] The fermentation media used according to the invention for culturing microorganisms usually also comprise other growth factors such as vitamins or growth pro-

moters, which include, for example, biotin, riboflavin, thiamine, folic acid, nicotinic acid, panthothenate, and pyridoxine. Growth factors and salts are frequently derived from complex media components such as yeast extract, molasses, cornsteep liquor and the like. It is moreover possible to add suitable precursors to the culture medium. The exact composition of the media compounds heavily depends on the particular experiment and is decided upon individually for each specific case. Information on the optimization of media can be found in the textbook "Applied Microbiol. Physiology, A Practical Approach" (Editors P.M. Rhodes, P.F. Stanbury, IRL Press (1997) pp. 53-73, ISBN 0 19 963577 3). Growth media can also be obtained from commercial suppliers, for example Standard 1 (Merck) or BHI (brain heart infusion, DIFCO) and the like.

[0209] All media components are sterilized, either by heat (20 min at 1.5 bar and 121° C.) or by filter sterilization. The components may be sterilized either together or, if required, separately. All media components may be present at the start of the cultivation or added continuously or batchwise, as desired.

[0210] The culture temperature is normally between 15° C. and 45° C., preferably at from 25° C. to 40° C. and may be kept constant or may be altered during the experiment. The pH of the medium should be in the range from 5 to 8.5, preferably around 7.0. The pH for cultivation can be controlled during cultivation by adding basic compounds such as sodium hydroxide, potassium hydroxide, ammonia and aqueous ammonia or acidic compounds such as phosphoric acid or sulfuric acid. Foaming can be controlled by employing antifoams such as, for example, fatty acid polyglycol esters. To maintain the stability of plasmids it is possible to add to the medium suitable substances having a selective effect, for example antibiotics. Aerobic conditions are maintained by introducing oxygen or oxygen-containing gas mixtures such as, for example, ambient air into the culture. The temperature of the culture is normally 20° C. to 45° C. and preferably 25° C. to 40° C. The culture is continued until formation of the desired product is at a maximum. This aim is normally achieved within 10 to 160 hours.

[0211] The fermentation broth can then be processed further. The biomass may, according to requirement, be removed completely or partially from the fermentation broth by separation methods such as, for example, centrifugation, filtration, decanting or a combination of these methods or be left completely in said broth. It is advantageous to process the biomass after its separation.

[0212] According to an aspect of the invention there is provided a method for the manufacture of a β-etherase polypeptide comprising the following steps:

[0213] i) provide a cell according to the invention and cell culture medium,

[0214] ii) culture the host cell in i) above to express the polypeptide according to the invention; and optionally,

[0215] iii) isolating said polypeptide from the cell or cell culture medium.

[0216] In a preferred method of the invention said cell is a microbial cell.

[0217] Preferably, said microbial cell is a bacterial or fungal host cell.

[0218] Protocols for the manufacture of recombinantly expressed proteins are known to the skilled person. Isolating proteins under denaturing conditions can result in a higher yield of the protein of interest when compared to non-denaturing protein purification methods. The purified denatured proteins are subsequently allowed to re-fold into their native structure.

[0219] In a further method said polypeptide isolation is under denaturing conditions.

[0220] According to an aspect of the invention there is provided a composition comprising or consisting of one or more polypeptides according to the invention.

[0221] In a preferred embodiment of the invention said composition comprises at least the polypeptide is set forth in SEQ ID NO:9 or 26

[0222] In a further preferred embodiment of the invention said one more polypeptide is set forth in SEQ ID NO: 9, 10, 11, 12, 13, 14, 15 and 16.

[0223] In a further preferred embodiment of the invention said one more polypeptide is set forth in SEQ ID NO: 26, 27, 28, 29, 30, 31, 32 and 33.

[0224] In a further preferred embodiment of the invention said composition further comprises one or more polypeptides for the saccharification of lignocellulose selected from the group consisting of cellulases, lytic polysaccharide monooxygenases, carbohydrate esterases, hemicellulases, glycosylhydrolases, endoglucanases, cellobiohydrolases, beta-glucosidases, xylanases, mannases, cellobiose dehydrogenases, and beta-xylosidases.

[0225] Saccharification is the process of breaking down complex carbohydrates such as cellulose into polysaccharides, disaccharides, and monosaccharides.

[0226] In a further preferred embodiment of the invention said composition comprises a buffer.

[0227] In a preferred embodiment of the invention said composition has a pH between 5 and 12, more preferably between 6 and 11, even more preferably between 7 and 10.

[0228] In a preferred embodiment of the invention said composition has a pH of 10.

[0229] In a preferred embodiment of the invention said composition has a pH of 7.

[0230] According to an aspect of the invention there is provided a method for the modification of plant biomass comprising the following steps:

[0231] I) contacting plant biomass with a composition or cell according to the invention to form a reaction mixture and

[0232] II) incubating said reaction mixture under conditions which cleaves β-ether linkages present the plant biomass to obtain depolymerised lignin units.

[0233] Plant biomass in the context of this application comprises or consist of lignin and/or lignocellulose.

[0234] In a preferred method of the invention said method comprises further step iii) extracting said depolymerised lignin units from the reaction mixture.

[0235] In a preferred method of the invention said depolymerised lignin units are selected from the group consisting of flavones, p-coumaric acid, and ferulic acid.

[0236] In a further preferred method of the invention said depolymerised lignin units are selected from the group consisting of flavones and p-coumaric acid.

[0237] In a further preferred method of the invention said depolymerised lignin units are selected from the group consisting of flavones, monomeric guaiacyl phenylpropanoid units, monomeric syringyl phenylpropanoid units, and monomeric p-hydroxyphenyl phenylpropanoid units.

[0238] In a further preferred method of the invention said flavones are tricin.

[0239] In a further preferred method of the invention said depolymerised lignin units are tricin and/or p-coumaric acid.

[0240] In a further preferred method of the invention said plant biomass is selected from hardwood and softwood or woody biomass.

[0241] In the context of this application woody biomass defines saw mill or paper mill discards.

[0242] In a further preferred method of the invention said plant biomass is selected from grasses, corn stover, corncob, corn fiber, wheat straw, sugarcane bagasse, wood pulp, rice straw, and municipal solid waste.

[0243] In a further preferred method of the invention said plant biomass is wheat straw or sugarcane bagasse.

[0244] In a further preferred method of the invention said method comprises further step of contacting the reaction mixture of iii) with a saccharification composition comprising one or more polypeptides for the saccharification of depolymerised lignin units.

[0245] In a preferred further method of the invention said saccharification composition comprises or consist of one or more polypeptides selected from the group consisting of cellulases, lytic polysaccharide monooxygenases, carbohydrate esterases, hemicellulases, glycosylhydrolases, endoglucanases, cellobiohydrolases, beta-glucosidases, xylanases, mannases, cellobiose dehydrogenases, and beta-xylosidases

[0246] In an alternative preferred method of the invention said saccharification composition is provided during step i).

[0247] In a preferred method of the invention said method comprises extracting di- and/or monosaccharides.

[0248] In a preferred method of the invention said monosaccharides are selected from the group consisting of glucose, xylose, and arabinose

[0249] According to an aspect of the invention there is provided the use of the polypeptides, cells or composition according to the invention in the hydrolysis of lignocellulose.

[0250] According to a further aspect of the invention there is provided a bioreactor comprising a cell or composition according to the invention.

[0251] In a preferred embodiment of the invention said bioreactor is a fermenter.

[0252] Throughout the description and claims of this specification, the words "comprise" and "contain" and variations of the words, for example "comprising" and "comprises", means "including but not limited to", and is not intended to (and does not) exclude other moieties, additives, components, integers or steps. "Consisting essentially" means having the essential integers but including integers which do not materially affect the function of the essential integers.

[0253] Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used, the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

[0254] Features, integers, characteristics, compounds, chemical moieties or groups described in conjunction with an aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

[0255] An embodiment of the invention will now be described by example only and with reference to the following figures:

[0256] FIG. **1**. Composition of prokaryotic and eukaryotic genera during wheat straw degradation. Sequences were generated on an ion torrent platform after amplification of the 16S and ITS for a) prokaryotic and b) eukaryotic identification, respectively. Operational taxonomic units were identified to genus level N=1;

[0257] FIG. **2**. Expression change of contigs between glucose and wheat straw conditions. RNA was extracted and sequenced after a) two, b) four and c) ten days of *P. putredinis* NO1. incubation on wheat straw and four days of growth on glucose. Points represent the log fold change (FC) and average counts per million (CPM) of contigs, between the wheat straw and glucose conditions. Carbohydrate-active enzymes were annotated using dbCAN namely auxiliary activities (AA), glycoside hydrolases (GH), polysaccharide lyases (PL), carbohydrate esterases (CE), glycosyltransferases (GT), and non-catalytic carbohydrate-binding modules (CB). Points are the average of three biological replicates;

[0258] FIG. **3**. Molar percentages of supernatant (SNT) and biotin-labelled (BF) proteins after four days of incubation on wheat straw. Molar percentages of carbohydrate-active families, GH: Glycoside hydrolase, AA: Auxiliary activity, PL: Polysaccharide lyase, CE: Carbohydrate esterase, and GT glycosyl transferase, were calculated as the sum of contigs annotated and taken as an average for each biological replicate. N=3;

[0259] FIG. **4**. Release of compounds after incubation with lignocellulosic biomasses. Biomass was treated for 16 h with our recombinant β-etherase, mushroom tyrosinase, and buffer alone, and reaction products were extracted with ethyl acetate, a) Tricin 1 release from wheat straw was observed and compared to an authentic standard using a High-Performance Liquid-Chromatography (HPLC), and mass was confirmed by time-of-flight mass spectrometry. b) HPLC analysis of enzyme incubations with sugarcane bagasse. Products were identified by mass spectrometry and comparison with authentic standards, as p-hydroxybenzaldehyde 2, vanillin 3, p-coumaric acid 4;

[0260] FIG. **5**. Release of sugars from sugarcane bagasse, wheat straw, and rice straw. Sugarcane bagasse, wheat straw, and rice straw were treated with recombinant β-etherase, commercial mushroom tyrosinase, and buffer only for 16 h prior to the application of Celluclast® commercial saccharification cocktail. Sugar release was calculated from the reaction mixture using High-Performance Anion-Exchange chromatography. Error bars represent the standard deviation of five biological replicates;

[0261] FIG. **6**. Optimisation of *P. putredinis* NO1 growth media. a) A central composite design was used to create a response surface morphology to yeast extract and sodium nitrate concentrations. b) Both cellulase and xylanase production was improved with a high yeast extract and low nitrate concentrations;

[0262] FIG. **7**. Growth of *P. putredinis* NO1 on wheat straw over a period of one month. a) Growth of *P. putredinis* NO1 on wheat straw estimated by the total protein present in the culture and b) the dried weight of the total biomass within the culture. c) The pH of the culture was also monitored alongside d the release of sugar after 1 h from 10%

supernatant loading on carboxymethylcellulose and beechwood xylan;

[0263] FIG. **8**. Proteomics of *P. putredinis* NO1 grown on wheat straw. a) Total proteins recovered from *P. putredinis* NO1 exoproteome across timepoints. b) Total molar percentage of CAZy class across timepoints in the biotin labelled protein sample and supernatant;

[0264] FIG. **9**. GGβ4MU β-etherase assay. Under the action of a β-etherase the 4—O—β-ether linkage is cleaved liberating the product MUF. Upon excitement at 372 nm MUF will fluoresce at 445 nm;

[0265] FIG. **10**. c2092_g1_i1 abundance within the a) transcriptomic and b) proteomic libraries. Circles represent sample values of biological replicates (N=3), and error bars ± SD of the mean;

[0266] FIG. **11**. Alignment of β-etherase amino acid sequence (c2092) with structurally related enzymes. Alignment with 2Y9W; tyrosinase from *Agaricus bisporus* (common mushroom), 2P3X; *Vitis vinifera* Polyphenol Oxidase, 4J3P; catechol oxidase *Aspergillus oryzae*, 1WX2; *Streptomyces castaneoglobisporus tyrosinase*, 4J6V; *Bacillus megaterium* N205D tyrosinase. Identical amino acids are indicated by asterisks and amino acids similarity by dots. The conserved N-terminal arginine residue is circled ; copper-binding regions are highlighted;

[0267] FIG. **12**. Reads per kilobase per million (RPKM) of contigs identified as sharing significant similarity of the putative β-etherase. Reads with a similarity identity of over 30% to c2092 were considered as displaying significant homology. Circles represent sample values of biological replicates (N=3), and error bars ± SD of the mean;

[0268] FIG. **13**. Activity of the putative β-etherase against the synthetic substrate GGβ4MU. a) Fluorescence activity of purified β-etherase against tyrosinase and buffer control reaction. b-c) optimum temperature and pH for purified β-etherase as assessed by GGβ4MU assay. Circles represent sample values, and bars sample mean ± SD, N=3;

[0269] FIG. **14**. UV spectrum showing oxidase activity of β-etherase against tyrosinase substrates. Either was incubated in 50 mM Tris pH 8.5 at room temperature with 1 mM of substrate against enzyme only or substrate only as controls, a) $_L$-DOPA reaction with tyrosinase, b) $_L$-DOPA reaction with β-etherase, c) tyrosine reaction with tyrosinase, d) tyrosine reaction with β-etherase;

[0270] FIG. **15**. UV spectrum showing oxidase activity of β-etherase against different phenolic compounds. 1 mg/mL of the enzyme was incubated in 50 mM Tris pH 8.5 at room temperature with 1 mM of either catechin hydrate, pyrogallol, vanillic acid, p-hydroxybenzoic acid or quercetin. UV-Vis spectra were recorded at regular intervals; and

[0271] FIG. **16**. Release of products from lignocellulosic substrates after incubation with β-etherase, mushroom tyrosinase and buffer only. Reactions were performed at physiological -pH 8.5 & 30° C. prior to the reaction products being extracted from the reaction supernatant using ethyl acetate and analysed with high-performance liquid-chromatography. Circles represent the individual sample values (N=5), and error bars ± SD of the mean.

[0272] FIG. **17**. Lignin aromatic and side-chain region of 2D HSQC NMR spectra (DMSO-d_6:pyridine-d_5, 4:1, v/v) of enzyme lignins (EL) from (A) the wheat control, and (B) the enzyme-treated wheat. Signal assignments in the spectra correspond to the chemical structures of the lignin monomeric subunits shown (S) syringyl, (G) guaiacyl, (H) p-hydroxyphenyl, (T) tricin, (pCA) p-Coumarate, (A) β-aryl ether (β—O—4), (B) phenylcoumaran (β-5), (C) resinol (β-β).

[0273] The quantification values shown in the table are for relative comparisons of the lignin components determined from NMR contour volume-integrals based on S + G + H = 100%. The pCA and T units are lignin appendages; their levels were estimated and expressed based on the total lignin (S + G + H). Assignments are from papers noted in the Experimental Section, along with the newly Aβ-T assignment (80). Note that, to allow the crucial lignin side-chain contours to be more clearly seen, the boxed lignin side-chain region was vertically scaled by ~1.75×.

[0274] FIG. **18**. SDS-PAGE after denaturation, purification and refolding. L is protein marker -Thermo Scientific™ PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa. E1 is protein purified in the absence of $CuSO_4$, and E2 was purified with $CuSO_4$ present in the refolding buffer.

TABLE 1

Proteins showing homology to the putative β-etherase within *P. putredinis* NO1 transcriptome. BLASTp searches were performed on the c2092_g1_i1 sequence (SEQ ID NO 9) against the assembled *P. putredinis* NO1 transcriptome

SEQ ID	evalue	pident	length	bitscore	Similarity%	Similarity
c19124_g1_i1_4 (SEQ IQ NO 10)	9.4E-111	43.796	411	330	0.608	256/421
c7740_g1_i1_6 (SEQ ID NO 11)	8.17E-77	38.482	382	243	0.508	23/439
c10688_g1_i1_2 (SEQ ID NO 12)	1.72E-74	40.395	354	236	0.52	226/435
c5294_g1_i1_3 (SEQ ID NO 13)	1.65E-71	37.366	372	229	0.52	223/429
c2117_g1_i1_2 (SEQ ID NO 14)	2.9E-57	36.936	349	191	0.422	184/436
c19010_g1_i1_4 (SEQ ID NO 15)	2.94E-32	29.254	335	125	0.325	164/505
c7470_g1_i1_2 (SEQ ID NO 16)	2.25E-26	23.37	368	108	0.376	169/449

TABLE 2

Proteins with homology to the β-etherase within NCBI non-redundant database. BLASTp searches were performed on the c2092_g1_i1 sequence against the non-redundant protein database held by NCBI. Results were filtered to >50 % identity

	Description	Max Score	Total Score	Query Cover	E value	Percent identity
gb|PKS12997.1|	hypothetical protein jhhlp_000338 [Lomentospora prolificans]	713	713	100%	0.0	87.50%
ref|XP_016642676.1|	Tyrosinase central domain protein [Scedosporium apiospermum]	674	674	100%	0.0	82.40%
gb|TPX10091.1|	hypothetical protein E0L32_001288 [Phialemoniopsis curvata]	572	572	93%	0.0	67.19%
gb|ELA32929.1|	tyrosinase central domain protein [Colletotrichum fructicola Nara gc5]	506	506	99%	7e-176	57.95%
gb|KZL67883.1|	tyrosinase central domain-containing protein [Colletotrichum tofieldiae]	501	501	97%	8e-174	58.90%
gb|EQB58959.1|	hypothetical protein CGLO_00722 [Colletotrichum gloeosporioides Cg-14]	497	497	92%	3e-172	59.89%
gb|KZL82263.1|	tyrosinase central domain-containing protein [Colletotrichum incamum]	496	496	97%	3e-172	58.15%
gb|KXH49404.1|	tyrosinase central domain-containing protein [Colletotrichum nymphaeae SA-01]	486	486	99%	2e-168	55.88%
gb|KXH49404.1|	tyrosinase central domain-containing protein [Colletotrichum simmondsii]	485	485	99%	1e-167	55.64%
gb|OLN85731.1|	Grixazone synthase 2 [Colletotrichum chlorophyti]	484	484	92%	3e-167	58.99%
ref|XP_018157362.1|	tyrosinase central domain-containing protein [Colletotrichum higginsiamamIMI 349063]	481	481	92%	4e-166	59.37%
gb|EXF76797.1|	tyrosinase central domain-containing protein [Colletotrichum fioriniae PJ7]	479	479	99%	2e-165	55.15%
gb|TDZ75107.1	tyrosinase-like protein orsC {colletotrichum trifolii]	476	476	92%	4e-164	59.95%
gb|TKW48599.1|	hypothetical protein CTA1_467 [Colletotrichum tanaceti]	473	473	92%	7e-163	58.42%
gb|TDZ15437.1|	tyrosinase-like protein orsC [colletotrichum orbiculare MAFF 240422]	470	470	92%	4e-162	60.48%
ref|XP_001227696.2|	hypothetical protein CHGG 09769 [Chaetomium globosum CBS 148.51]	469	469	100%	2e-161	55.50%
gb|TDZ29471.1|	Tyrosinase-like protein orsC [colletotrichum spinosum]	460	460	92%	2e-157	57.00%
ref|XP_022470530.1|	tyrosinase central domain-containing protein [Colletotrichum orchidophilum]	458	458	99%	2e-157	54.66%
gb|OIW32989.1	tyrosinase central domain-containing protein [Coniochaeta ligniaria NRRL30616	447	447	92%	5e-153	53.79%
gb|KXH30586.1|	tyrosinase central domain-containing protein [Colletotrichum salicis]	447	447	97%	3e-152	54.02%
gb|RKU41032.1|	hypothetical potein DL546 002981 [Coniochaeta pulveracea]	442	442	99%	5e-151	51.96%
gb|KZL64229.1|	tyrosinase central domain-containing protein [Colletotrichum incanum]	434	434	92%	4e-145	55.17%
gb|TEA15757.1|	Tyrosinase-like protein orsC [Colletotrichum sidae]	427	427	92%	6e-145	55.00%
gb|OHW92206.1|	tyrosinase central domain-containing protein [Colletotrichum incanum]	420	420	84%	5e-143	57.73%
ref|XP_01816298.1|	Tyrosinase central domain-containing protein [Colletotrichum higginsianum IMI 349063]	425	425	92%	1e-142	54.38%
gb|TID02585.1|	Tyrosinase ustQ [Colletotrichum higginsianum]	425	425	92%	1e-142	54.38%

TABLE 2-continued

Proteins with homology to the β-etherase within NCBI non-redundant database. BLASTp searches were performed on the c2092_g1_i1 sequence against the non-redundant protein database held by NCBI. Results were filtered to >50 % identity

	Description	Max Score	Total Score	Query Cover	E value	Percent identity
gb|OLN83361.1|	Tyrosinase 2 [Colletotrichum chlorophyti]	417	417	92%	5e-141	51.97%
emb|CCF32411.1|	hypothetical protein CH063 04807 [Colletotrichum higginsianum	412	412	84%	7e-140	56.85%
gb|KZL72889.1|	tyrosinase-like protein [Colletotrichum tofieldiae]	412	412	84%	7e-140	57.14%
gb|TKW50870.1|	hypothetical protein CTA1 3684 [Colletotrichum tanaceti]	419	419	92%	7e-140	52.39%
gb|KDN70624.1|	hypothetical protein CSUB01 04485 [Colletotrichum sublineola]	417	417	92%	1e-139	53.58%
gb|EXF84421.1|	hypothetical protein CFIO01_02736 [Colletotrichum fioriniae PJ7]	409	409	92%	1e-136	52.22%
gb|XP_003664995.1|	tyrosinase-like protein [Thermothelmyces thermophilus ATCC 42464]	404	404	92%	3e-136	54.09%
gb|TQN72542.1	Tyrosinase-like protein orsC [Colletotrichum sp. PG-2018a]	407	407	89%	5e-136	54.77%
ref|XP_003351009.1|	uncharacterized protein SMAC 04313 [Sordaria marcrospora k-hell]	399	399	97%	6e-134	50.12%
ref|XP_006692366.1|	hypothetical protein CTHT 0018720 [Chaetomium thermophilum yar. thermophilum DSM 1495]	395	395	89%	1e-132	54.67%
gb|TDZ58291.1|	Tyrosinase-like protein orsC [Colletotrichum trifolii]	393	393	79%	6e-132	57.67%
gb|TDZ23501.1|	Nitroalkane oxidase [Colletotrichum orbiculare MAFF 240422]	409	409	80%	8e-132	57.75%
ref|XP_022471338.1|	hypothetical protein COR01 10513 [Colletotrichum orchidophilum]	397	397	92%	9e-132	50.78%
gb|KXH34366.1|	hypothetical protein CSIM01 00277 [Colletotrichum simmondsii]	396	396	92%	2e-131	50.51%
gb|KXH69104.1|	hypothetical protein CSAL01 01466 [Colletotrichum salicis]	389	389	81%	3e-129	56.19%
ref|XP_008090963.1|	hypothetical protein GLRG 02114 [Colletotrichum graminicola M1.001	378	378	79%	2e-126	56.44%
ref|XP_001227853.1|	hypothetical protein CHGG 09926 [Chaetomium globosum CBS 148.51]	373	373	92%	5e-124	50.00%
gb|TDZ28941.1|	Tyrosinase-like protein orsC [Colletotrichum spinosum]	371	371	73%	2e-122	58.14%
gb|ELA37064.1|	hypothetical protein CGGC5 3508 [Colletotrichum fructicola Nara gc5]	364	364	72%	1e-121	59.52%
ref|XP_007911158.1	putative tyrosinase-like protein [Phaeoacremonium minimum UCRPA7]	363	363	68%	2e-121	59.22%
gb|EQB52888.1|	hypothetical protein CGLO 07432 [Colletotrichum gloeosporioides Cg-14]	361	361	72%	2e-120	59.86%
gb|TEA10724.1|	Nitroalkane oxidase [Colletotrichum sidae]	373	373	73%	4e-118	58.33%
ref|XP_024731024.1|	putative tyrosinase [Meliniomyces bicolor E]	331	331	79%	2e-108	51.38%
emb|CDP29730.1|	Putative tyrosinase [Podospora anserina S mat+	326	326	81%	4e-106	50.15%
emb|VBB81548.1|	Putative tyrosinase [Podospora comtat]	326	326	81%	5e-106	50.15%
ref|XP_001273822.1|	tyrosinase, putative [Aspergillus clavatus NRRL 1]	326	326	83%	2e-105	50.00%
ref|XP_001905273.1|	uncharacterized protein PODANS 5 7820 [Podospora anserina S mat+]	323	232	80%	3e-105	50.00%
gb|PGH18781.1|	hypothetical protein AJ79_00194 [Helicocarpus griseus UAMH5409]	325	325	83%	5e-105	50.15%
gb|PBP21500.1|	hypothetical protein BUE80 DR007716 [Diplocarpon rosae]	278	278	68%	4e-88	50.17%

TABLE 3

Purification of β-etherase. The heterologously expressed protein was purified using anion-exchange (Q) and size-exclusion chromatography (S.E). Protein concentration and VT221 activity was calculated after each purification step

Purification steps	Total Protein mg	Activity (mU) (nmol/mg/hr)	Specific (U/mg)	Yield (%)	Purification fold
Culture filrate	1024	7500	7.32	100	1
Q	29.25	2600	88	34.67	12
S.E	14	1950	139	26	19

TABLE 4

β-etherase substrate specificity

Substrate	Etherase reactivity	Tyrosinase reactivity
Tyrosine methyl ester	—	+
L-Dopa (3,4-dihydroxy-L-phenylalanine)	—	+
Dopamine hydrochloride	—	+
Caffeic acid (catechol oxidase substrate)	—	+
4-Methly-catechol (catechol oxidase substrate)	—	+
Tyrosol (catechol oxidase substrate)	—	—
Tannic acid	—	—
(+)-Catechin hydrate	+	+
Pyrogallol	+	+
4-Hydroxybenzoic acid	+	—
Quercetin	+	—
Vanillic acid	+	—

MATERIAL AND METHODS

Wheat Straw Degradation in Shake-Flasks Inoculated with Compost

[0275] Two-liter shake flasks, containing 1 L minimal media and 5% (w/v) milled wheat straw, were inoculated with 1% (w/v) compost. The inoculum was collected from composting wheat straw that had been developed over the period of a year and watered at regular intervals. The inoculum was prepared by blending until homogenized and used on the day of preparation. The minimal media was based on *Aspergillus niger* minimal media and contained KCI 0.52 g/L, KH_2PO_4 0.815 g/L, K_2HPO_4 1.045 g/L, $MgSO_4$ 1.35 g/L, $NaNO_3$ 1.75 g/L, Hutner's trace elements ($Na_2EDTA \cdot 2H_2O$ 50 g/L, $ZnSO_4 \cdot 7H_2O$ 22 g/L, H_3BO_3 11.4 g/L, $MnCl_2 \cdot 4H_2O$ 0.506 g/L, $FeSO_4 \cdot 7H_2O$ 0.4499 g/L, $CoCl_2 \cdot 6H_2O$ 0.161 g/L, $CuSO_4.5H_2O$ 0.157 g/L, $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ 0.110 g/L). These flasks were incubated at 30° C. and shaken at 150 rpm. Aliquots (10 mL) containing both the solid and liquid fractions were aseptically collected weekly for eight weeks. The samples were then serially diluted with x1 phosphate-buffered saline to concentrations ranging between 10^{-1} and 10^{-7}. From these dilutions 100 μL samples were used to create spread plates on both nutrient agar (NA) and potato dextrose agar (PDA), in order to culture strains from the composting environment.

Targeted Amplicon Sequencing of 16S and ITS Region

[0276] Genomic DNA was harvested from the compost cultures using a modified CTAB protocol adapted for use on materials with high phenolic contents. From the composting shake flask, 20 mL aliquots were harvested weekly. The biomass was separated from the liquid fraction by centrifugation performed at 4000 g at 4° C., and 0.5 g of biomass removed to a 2 mL screw-cap tube. To this 500 μL of cetyltrimethylammonium bromide (CTAB) buffer (2% (w/v) CTAB 100 mM Tris-HCI (pH 8.0), 20 mM EDTA (pH 8.0), 2 M NaCl, 2% (w/v) polyvinylpyrrolidone (Mr 40.000), 5% 2-mercaptoethanol (v/v), 10 mM ammonium acetate, was added along with 0.5 g of zirconia beads and 0.5 mL of phenol: chloroform: isoamyl alcohol (25: 24: 1, pH 8.0), before briefly vortexing. The material was then bead-beaten using a TissueLyser II (Qiagen) for 5 min at speed 28/s. A modified phenol-chloroform method was used to extract DNA after cell lysis. The sample was spun for 5 min at max speed to achieve separation of the phases before the aqueous layer was removed to a fresh 2 mL Eppendorf tube. To the aqueous phase chloroform: isoamyl alcohol (21:1) was added, and this was spun and the aqueous phase transferred to a fresh tube, to remove any remaining phenolics. To precipitate the DNA within the sample, an equal volume of ice-cold 100% isopropanol was added and incubated for 1 h. DNA was pelleted by centrifugation at 13,000 rpm for 10 min, and supernatant was removed without disturbing the pellet. The pellet was then washed with 80% ethanol, before being resuspended in DNAse-free water.

[0277] Regions for amplicon sequencing were amplified using Phusion® High-Fidelity DNA Polymerase (Finnzymes OY, Finland) as per manufactures instructions before being purified with Agencourt AMPure XP (Beckman Coulter), and sequenced at the Biorenewables Development Centre (BDC), York, U.K. using an Ion Torrent platform. The primers pairs, for ITS and 16S sequencing, were as follows; ITS1 Fw - TCCGTAGGTGAACCTGCGG (SEQ ID NO 34), Rv - CGCTGCGTTCTTCATCG (SEQ ID NO 35), 16S Fw -AYTGGGYDTAAAGNG (SEQ ID NO 36), Rv-TACNVGGGTATCTAATCC(SEQ ID NO 37). Ribosomal DNA sequence data generated via targeted amplicon sequencing was analyzed using the open-access software Qiime on the University of York's Technology Facilities linux server. [57] Each fastq file generated from the IonTorrent platform was first demultiplexed and then converted into both fasta and qual file types using Qiimes python script convert_fastaqual_fastq.py. To remove the primer sequences from the reads, the script split_libraries.py was used along with a mapping file generated as per Qiimes instructions. Low-quality reads were removed by filtering out reads under 180 bp long and those without recognizable primers. The orientation of the sequences was then corrected based on the primer location. Operational taxonomic units (OTUs) were then created from the fasta files. These files were picked using the open-reference OTU picking process. To perform this, the script pick_open_reference_otus.py was used. This step also includes taxonomy assignment,

sequence alignment, and tree building steps. For the taxonomy assignments of bacterial sequences the default reference database was used, (greengenes gg_13_8 97_otus database),[58,59] and for the fungal ITS sequences the UNITE (alpha release 12_11) database was used.[60]

Central Composite Design for Media Optimisation

[0278] Media was optimized using a central composite design with rotation.[61] It was optimized for the production of both cellulase and xylanase enzymes after seven days on 1.5% wheat straw and minimal media, as assessed by measuring reducing sugar release after incubation on CMC and xylan. The concentrations of both sodium nitrate and yeast extract were varied as part of the optimization. The sodium nitrate concentration was varied between 0 g/L and 3.5 g/L, and yeast extract was varied between 0% and 1% (w/v). Statistica 6.0 software was used to create the experimental design and analyze the results.

[0279] The optimized media for *P. putredinis* NO1 growth consisted of yeast extract 8.55 g/L, KCI 0.52 g/L, KH_2PO_4 0.815 g/L, K_2HPO_4 1.045 g/L, $MgSO_4$ 1.35 g/L, $NaNO_3$ 1.75 g/L and Hutner's trace elements.

Characterization of *P. Putredinis* NO1 Growth on Wheat Straw

[0280] Growth of *P. putredinis* NO1 was assessed using the dried weight of the biomass present within the culture. Cultures were transferred to pre-weighed and freeze-dried falcon tubes and chilled for 5 min. They were then centrifuged at 4,500 rpm, and the supernatant removed. The biomass was gently rinsed with x1 PBS and tubes were flash-frozen in liquid nitrogen and lyophilized. Each tube was then re-weighed to calculate the dry weight of the biomass present. The total protein content of the cultures was used as an indicator of growth on insoluble materials such as wheat straw. Total protein was extracted by boiling 100 μg of freeze-dried biomass in 1 mL of 0.2% (w/v) sodium dodecyl sulfate, for 5 min to lyse all cells present. Protein was then collected by centrifugation at 14,000 rpm and the supernatant collected into a fresh 50 mL falcon tube. This was repeated three times, without heating, and with vigorous vortexing between each centrifuge step to wash the biomass of any remaining protein. Extracted protein was precipitated with five volumes of ice-cold acetone overnight at -20° C., before being centrifuged at 4500 rpm and the resulting pellet washed with 80% (v/v) ice-cold ethanol. The ethanol-protein mix was then centrifuged again, and the supernatant removed and the pellet air-dried. The protein was then solubilized in 3 mL of H_2O and quantified using the Bradford assay. The ability of an enzyme to cleave polysaccharides and produce products with reducing ends was assessed at each timepoint by incubating 10 μL of cultural supernatant with the 2% (w/v) of either carboxymethylcellulose (CMC) or xylan (beechwood) in 200 μL of 50 mM sodium phosphate at 6.8 and 30° C. Before and after incubation 10 μL aliquots mixed with p-hydroxybenzoic acid hydrazide (PAHBAH), heated to 70° C. for 10 min, and color change detected at 415 nm using a microtitre Tecan Safire2 plate reader.[62] A stock solution of the appropriate monosaccharide was assayed to obtain a standard curve for quantification of sugar release.

RNA Extraction from *P Putredinis* NO1 Sp

[0281] Cultures of *P. putredinis* NO1 were established in 200 mL shake flasks, containing 20 mL of the optimized growth media and either 1.5% wheat straw or 0.5% glucose. These were incubated at 30° C. with shaking at 180 rpm. To control for varying amounts of cell growth, aliquots of either 0.5 g, 0.3 g and 0.1 g of biomass from the wheat straw cultures were weighed into 2 mL screw-cap tubes that contained 3×3 mm tungsten carbide beads and 1 mL Trizol (Life Technologies). The cells were then disrupted in a TissueLyser II (Qiagen) for either 2×2 min or 2×5 min at 28/s, dependent on the stage of growth. Total RNA was then extracted with the standard Trizol method as per manufacturer's instructions and extracted RNA was resuspended in 50 μL of nuclease-free water. The quality of RNA was assessed by visualization on agarose gels. To obtain enough RNA for processing six technical replicates were performed for each biological replicate. These were stored at -80° C. after being flash-frozen in liquid nitrogen before further processing could occur. The RNA samples were treated for DNA contamination with RTS DNase kits (Mobio) using standard methods described by the manufacturers. The samples were then cleaned with ZymoResearch RNA Clean &

[0282] Concentrator™ 5 kits, using the manufacturer's protocol to separate small and large RNA fragments into different fractions. RNA fragments greater than 200 nt were elution into 50 μL of RNase-free water before RNA concentration, and quality was evaluated with the 2200 TapeStation (Aligent). Once total RNA of a suitable quantity and quality was obtained, samples could be enriched for messenger RNA (mRNA). This was performed using Ribo-Zero™ Magnetic Epidemiology rRNA removal kit (RZE1224/MRZ11124C; Illumina) according to the manufacturer's protocol.

RNA Sequencing

[0283] The Genome Analysis Centre (TGAC), Norwich, U.K, performed the RNA sequencing on an Illumina HiSeq platform. As per the requirements of the sequencing service, 100 ng of enriched mRNA was provided for each sample. From the proved mRNA, cDNA libraries were constructed using the adapted TruSeq RNA v2 protocol (Illumina 15026495 Rev.B). Libraries were then normalized using elution buffer (Qiagen) and pooled in equimolar amounts into one final 12 nM pool. These were then diluted to a final concentration of 10 pM, spiked with 1% PhiX and loaded onto the Illumina cBotTemplate, for hybridization and first extension, using the TruSeq Rapid PE Cluster Kit v1 before the flow cell was transferred onto the Illumina HiSeq2500. Here, the remainder of the clustering process was conducted, and the library pool was run in a single lane for 100 cycles of each paired-end read before samples were demultiplexed. One base-pair mismatch per library was allowed, and reads were converted to FASTq. The raw data was subject to rRNA removal by catching the remaining paired reads after mapping to a modified rRNA_115_tax_silva_v1.0 ribosomal set, using BOWTIE2. The reads were further trimmed to remove adaptor sequences with the ngsShoRT_2.1 method, and libraries were pooled before being assembled by Trinity Software to obtain 37,720 contigs. Then, using this assembly as a reference, the original (unprocessed) individual libraries were mapped and the

number of reads counted for each contig. Counts per million (CPM) were converted to reads per kilobase of exon per million reads mapped (RPKM) to normalize for both the depth of sequencing achieved in each sample and length of the contig.

[0284] Emboss GETORF (http://www.bioinformatics.nl/cgi-bin/emboss/getorf) was used to generate putative protein-coding sequences by translating all regions over 300 bp between potential start and stop codons. Putative open reading frames (ORFs) were searched against the NCBI non-redundant protein database and KOG database using BLASTp, and Pfam and dbCAN databases using HMMER3.(45, 81, 82) Local BLAST searches using unique were performed using BLAST+ 2.3.0.(65, 64) Signal peptides were predicted from ORFs using SignalP 4.0.(66, 67)

Protein Extraction

[0285] Supernatant proteins were harvested by collecting samples (20 mL) from the culture supernatant of *P. putredinis* NO1 and precipitated in five volumes of ice-cold acetone. The acetone fractions were incubated overnight at -20° C., before being centrifuged at 10,000 xg. The resulting pellet was washed with 80% ice-cold acetone, air-dried and resuspended in 0.5x PBS with 0.1% sodium dodecyl sulfate (SDS). To selectively extract biomass bound proteins, two grams of biomass collected from the fungal cultures was washed twice with ice-cold 0.5x PBS, before being resuspended and mixed for 1 h at 4° C., in 0.5x PBS with 10 mM EZ-linked biotin (Thermo Scientific). The reaction was then quenched for 30 min with 50 mM Tris-HCL, pH 8, and excess biotin was removed by washing twice with ice-cold 0.5 × PBS. Warmed SDS (2% w/v, at 60° C.) was used to extract the proteins. The mixture was incubated at room temperature for 1 h, centrifuged and precipitated with ice-cold acetone as described above. The resulting pellets were solubilized in 1x PBS containing 0.1% SDS then loaded onto streptavidin columns (Thermo Scientific) that had been pre-washed (0.1% SDS 1× PBS). The proteins were then incubated for 1 h on the column at 4° C., and washed with three column volumes of 0.1% SDS 1x PBS, before being incubated overnight with elution buffer (50 mM DTT in 1 × PBS) at 4° C. Proteins were eluted the following day by the addition of 1 mL elution buffer and the resulting fraction collected. The column was incubated for one hour before this was repeated. In total the elution was performed four times. These fractions were then flash-frozen in liquid nitrogen, freeze-dried, resuspended in 2 mL distilled water and desalted using Zeba, 7 K MWCO columns (Thermo Scientific) following manufacturer's instructions. Both the supernatant and biotin-tagged proteins were stored in 4-12% (w/v) Bis-Tris acrylamide gels. Protein samples were loaded into the gel, separated electrophoresis for 20 min and stained with InstantBlue (Sigma-Aldrich).

Proteomic LC-MS/MS

[0286] LC-MS/MS was performed to identify proteins within both the supernatant and biotin-labelled fractions. Proteins contained within gel slices were washed with 50% (v/v) aqueous acetonitrile containing 25 mM ammonium bicarbonate, then reduced with 10 mM DTE and S-carbamidomethylated with 50 mM iodoacetamide. Gels were then dehydrated with acetonitrile and digested with 0.2 μg trypsin (Promega) in 25 mM ammonium bicarbonate. The digestion was performed overnight at 37° C. Peptides were extracted with 50% (v/v) aqueous acetonitrile, dried in a vacuum concentrate and resuspended in 0.1% (v/v) aqueous trifluoroacetic acid. Peptides were loaded onto a nanoAcquity UPLC system (Waters) equipped with a nanoAcquity Symmetry C18, 5 μm trap (180 μm × 20 mm Waters) and a nanoAcquity HSS T3 1.8 μm C18 capillary column (75 mm × 250 mm, Waters). The trap was washed with 0.1% (v/v) aqueous formic acid at a flow rate of 10 μL min^{-1}, before switching to the capillary column. Peptides were separated using a gradient elution of two solvents, 0.1% (v/v) aqueous formic acid (solvent A) and acetonitrile containing 0.1% (v/v) formic acid (solvent B). The flow rate used was 300 nL min^{-1}, and the column temperature was 60° C. The gradient proceeded linearly from 2% solvent B to 30% over 125 min, then 30-50% over 5 min, before being washed with 95% solvent B for 2.5 min. The column was then re-equilibrated at the initial conditions for 25 min before subsequent injections. The nanoLC system was interfaced with a maXis HD LC-MS/MS System (Bruker Daltonics) with a CaptiveSpray ionization source (Bruker Daltonics). Positive ESI- MS & MS/MS spectra were acquired using AutoMSMS mode. Instrument control, data acquisition and processing were performed using Compass 1.7 software (microTOF control, Hystar and DataAnalysis, Bruker Daltonics). Instrument settings were as follows: ion spray voltage: 1,450 V; dry gas: 3 L min^{-1}; dry gas temperature 150° C.; collision RF: 1,400 Vpp; transfer time: 120 ms; ion acquisition range: m/z 150-2,000. AutoMSMS settings specified: absolute threshold 200 counts, preferred charge states: 2-4 , singly charged ions excluded. Cycle time: 1 s, MS spectra rate: 5 Hz, MS/MS spectra rate: 5 Hz at 2,500 cts increasing to 20 Hz at 250,000 cts or above. Collision energy and isolation width settings were automatically calculated using the AutoMSMS fragmentation table. A single MS/MS spectrum was acquired for each precursor, with dynamic exclusion for 0.8 min unless the precursor intensity increased fourfold.

Genomic Data Analysis

[0287] The raw data was subject to rRNA removal by catching the remaining paired reads after mapping to a modified rRNA_115_tax_silva_v1.0 ribosomal set, using BOWTIE2. The reads were further trimmed to remove adaptor sequences with the ngsShoRT_2.1 method, and libraries were pooled before being assembled by Trinity Software to obtain 37,720 contigs. Then, using this assembly as a reference, the original (unprocessed) individual libraries were mapped and the number of reads counted for each contig. Counts per million (CPM) were converted to reads per kilobase of exon per Million reads mapped (RPKM) to normalize for both the depth of sequencing achieved in each sample and length of the contig. Emboss GETORF (http://www.bioinformatics.nl/cgi-bin/emboss/getorf) was used to generate putative protein-coding sequences in all six reading frames from the transcriptomic libraries by translating regions over 300 bp long between potential start and stop codons. These putative open reading frames (ORFs) were searched against the NCBI non-redundant protein database and KOG database using BLASTp, the Pfam and dbCAN databases using HMMER3.[45,63] Annotations were subsequently mapped back to the contig from which the ORF originated. Local BLAST searches using unique

were performed using BLAST+ 2.3.0.[64,65] Signal peptides were predicted from ORFs using SignalP 4.0.[66,67]

Proteomic Data Analysis

[0288] Spectra obtained from the LC-MS/MS analysis were searched against all potential opening reads frames generated from the *P. putredinis* NO1 transcriptomic library, using Mascot (Matrix Science Ltd., version 2.4). This was locally run through the Bruker ProteinScape interface (version 2.1). Search criteria were specified as follows; the instrument was selected as ESI-QUAD-TOF, trypsin was stated as the digestion enzyme, fixed modifications as carbamidomethyl (C), and variable modifications as oxidation (M). Peptide tolerance was 10 ppm, and MS/MS tolerance 0.1 Da. Results were filtered through 'Mascot Percolator' to achieve a global false discovery rate of 1%, as assessed against a decoy database and further adjusted to accept only individual peptides with an expect score of 0.05 or lower. An estimation of relative protein abundance was performed as described by Ishihama,[68] whereby an exponentially modified Protein Abundance Index (emPAI) is used to estimate the relative abundance of proteins in LC-MS/MS experiments. From this index the molar percentage values could be calculated by normalising individual protein Mascot emPAI values against the sum of all emPAI values for each sample. Protein sequences were retrieved using the R package BioStrings.[69]

Synthesis of Synthetic Substrate GGβ4MU (7-[2-hydroxy-2-(4-hydroxy-3-methoxyphenyl)-1-(hydroxymethyl)ethoxy]-4-methyl-2H-1-benzopyran-2-one).

[0289] The synthetic substrate GGβ4MU was synthesized in 6 steps according to the protocol reported by Weinstein and Gold starting from acetovanillone.[44] The pure substrate GGβ4MU was obtained as a white solid following purification using plate chromatography on silica-gel (10% v/v MeOH in CH_2Cl_2). The NMR data were in excellent agreement with those previously reported.[44]

Identification of β-Etherase from Native Supernatant

[0290] *P. putredinis* NO1 was cultivated in medium containing 1.5% wheat straw. The supernatant was filtered, and the protein of interest purified by different purification steps, including ammonium sulfate precipitation (ASP), gel filtration using a superdex 200 (GF) on two different columns and anion-exchange chromatography (AE). Briefly, filtered culture supernatant with 0.1% Tween20 was concentrated in a 50 mL stirred Ultracentrifugation Cell (Millipore Corporation, USA) with a Biomax 30 kDa Ultrafiltration Membrane (Millipore Corporation, USA). Ammonium sulfate was slowly added to the filtered culture supernatant to a concentration of 20% while stirring at 4° C. The solution was centrifuged at 10.000 g for 15 min. The pellet was then resuspended in 2 mL buffer A (50 mM Tris-HCl, 100 mM NaCl, 0.1% Tween 20, pH 8.5). Additional ammonium sulfate was added to the supernatant, following the same procedure as described above, to obtain fractions with 30, 40 and 50% ammonium sulfate. After assessing the fractions with the GGβ4MU assay, samples were purified via gel-filtration on a Superdex-200 (GE Healthcare, US), using the ÄKTA system and 50 mM Tris-HCl, 100 mM NaCl, 0.1% Tween 20, pH 8.5. The most active sample was further purified using anion-exchange chromatography. Anion-exchange chromatography was conducted on a DEAE FF column (GE Healthcare, US) with an increasing salt concentration from 0 to 1 M NaCl in 20 min (5 mL/min). A running buffer of 30 mM Tris-HCl, 0.1% Tween 20, at various pH (7.0/7.4/8.5) was used. The Elution buffer was 30 mM Tris-HCl, 1 M NaCl, 0.1% Tween 20.

Gene Cloning and Expression

[0291] The c2092 gene was codon-optimized for expression in E. coli and synthesized into pET151 vector with N-terminal His-tag by Invitrogen. The expression plasmid was transformed into Arctic Express (DE3) competent cells, and successful transformants were selected on LB media containing ampicillin (100 mg L^{-1}) and gentamycin (10 mg L^{-1}). Auto-induction media was used for protein production. Inoculated cultures were incubated at 30° C. with shaking at 180 rpm until an optical density of 0.6 at 600 nm was reached. Once a suitable cell density was reached flasks, the temperature was reduced to 11° C. for 48 h before harvesting.

Purification of Recombinant β-Etherase

[0292] Cell pellets were collected by centrifugation at 7000 rpm and 4° C. for 15 min, then suspended in 50 mL (50 mM Tris, 1 mm DTT, pH 8.5). Suspended pellets were then sonicated on ice for using a Misonix S-4000 sonicator at 70 kHz for 4 min, using a program of 3 s off followed by 7 s on. After centrifugation at 17,000 rpm for 45 min to remove cell debris, the protein was purified by anion-exchange chromatography facilitated by an ÄKTA purifier UPC10 with UNICORN 5.31 workstation. Briefly, clear supernatant was loaded onto a mono-Q anion-exchange chromatography HP column (5 mL, GE Healthcare) that had previously been equilibrated with 50 mm Tris, 100 mm NaCl, 10% glycerol pH 8. The protein was then eluted with an increasing NaCl gradient (0 to 1 M) for 100 min at a rate of 1 mL/min. Eluted fractions containing the protein of interest were pooled and concentrated using Millipore Vivaspin20 10 kDa (Sartorius). These were then injected into a superdex 75 (16/60) gel-filtration column (GE Healthcare) that had been equilibrated with 50 mM Tris, 150 mM NaCl, 10% glycerol pH 8.5. Fractions were assessed with SDS-PAGE to determine purity, and the protein concentration was calculated spectroscopically using the extinction coefficient at 280 nm.

Purification and Refolding of Recombinant β-Etherase

[0293] Cell cultures were pelleted through centrifugation. Supernatant was discarded, and pellets were suspended in 5 mL per 100 mL of starting culture 20 mM (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (HEPES) pH 8, before sonicated on ice (70 V, 4 s on, 7 s off for a total of 4 min on). Centrifugation at 10 000 xg was again used to pellet cell debris and inclusion bodies. The pellet was washed with 20 mM HEPES, 2 M Urea, 0.5 M NaCl, 2% TritonTM X-100, pH 8, using the same volume as before, and sonicated and centrifuged as before. The resultant pellet was then resuspended in 20 mM HEPES, 0.5 M NaCl, 5 mM imidazole, 6 M guanidine hydrochloride, 1 mM dithiothrei-

tol (DTT) pH 8, using 10 mL per 100 mL of original cell culture, to solubilise inclusion bodies. After pelleting through centrifugation for a final time, the supernatant was applied to a HisTrap column equilibrated with 20 mM HEPES, 0.5 M NaCl, 5 mM imidazole, 6 M guanidine hydrochloride, 1 mM DTT pH 8. The equilibration buffer was then used to wash the column for a total of 5 CV followed by the same volume of 20 mM HEPES, 0.5 M NaCl, 20 mM imidazole, 6 M urea, 1 mM DTT pH 8. A linear gradient from the final wash buffer to 20 mM HEPES, 0.5 M NaCl, 20 mM imidazole, 0.1 mM $CuSO_4$, 1 mM DTT pH 8 was then used to refold the tagged protein on the column. This was applied over 30 mL using a flow rate of 0.5 ml/min. To elute refolded protein another linear gradient was applied over 20 mL, starting with 20 mM HEPES, 0.3 M $MgCl_2$, 20 mM imidazole, 1 mM DTT, pH 8 and ending with the same buffer with the addition of 500 mM imidazole and 10% glycerol. Apart from when otherwise mentioned, the flow rate was kept at 1 mL/ min when using a 1 mL capacity column and 3 mL/min when using a 5 mL capacity column. Fractions of 1.5 mL were collected throughout the elution step, and UV absorbance was used to determine protein content. Fractions with high protein contents were visualised using SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and the presence of the recombinant protein confirmed through western blot analysis. Protein activity was confirmed through the measurement of 4MU from the GGβ4MU assay after removal of imidazole and DTT using Zeba™ Spin Desalting Columns, 7K MWCO (ThermoFisher) or Slide-A-Lyzer™ Dialysis Cassettes 10 K MWCO (ThermoFisher).

Fluorescence Assay for β-Etherase

[0294] Enzyme activity was measured in 1 mL reaction containing 10 μL 4MU/GGβ4MU (synthetic fluorescent substrate 10 mM) and appropriate concentration of pure protein in 50 mM Tris-HCI, 100 mM NaCl, pH 8.5, 5 mM $CuSO_4$. The reaction was incubated at 30° C. for 1 h. Formation of 4-methylumbelliferone (4MU) was monitored using an RF-1500 fluorometric analyzer. After 0 h and 1 h of incubation 100 μL of the reaction mixture was taken and added to 50 μL of 100 mM glycine-NaOH buffer (pH 10.0). One unit of the enzyme was defined as the amount that released 1 nmol of 4 MU/h from the substrate. Five replicate were taken for each sample, and control reactions of boiled enzyme and wheat straw treated with buffer only were also performed.

Enzyme Properties

[0295] The effect of pH and temperature on enzyme activity was investigated by varying the pH of the reaction mixtures using 50 mM Tris-HCI buffer from pH 7.0 to 9.5, 50 mM glycine-NaOH buffer at pH range 9.0 to 10.5 and 50 mM Na_2HPO_4-NaOH buffer at pH range 10.5 to 12. The optimum temperature of enzyme activity was determined at various temperatures ranging from 20° C. to 70° C. Assays were performed as described in the previous section.

Phenol Oxidase Assay

[0296] Specificity was investigated by incubating 1 mM of each substrate of interest with the enzyme in 100 μL Tris pH 8.5 buffer at room temperature. Activity was determined by monitoring the change in Ultraviolet-Visible absorbance spectra (220 - 750 nm) of aliquots using a NanoDrop 8000 Microvolume UV-Vis spectrophotometer (Thermo Scientific). Scans were performed at regular intervals over 2 h.

Extraction of Tricin

[0297] Wheat straw was ground to <1 mm using a cyclone mill (Retsch) and washed several times with 50 mM Tris pH 8 to remove residual surface sugars. In 1 mL reactions, 100 mg of washed wheat straw was incubated with an appropriate concentration of pure enzyme in 50 mM Tris buffer at pH 8 with 5 mM $CuSO_4$. Reactions were incubated overnight at 30° C. with shaking. Control reactions were performed using wheat straw incubated with boiled β-etherase or with buffer only. Tricin was extracted based upon Karambelkar.[70] Briefly, 1 mL of ethyl acetate was added to 100 μL of the reaction supernatant. This was homogenized before being centrifuged for 5 min at 13,000 rpm. The ethyl acetate layer was transferred into new tubes and evaporated using a centrifugal evaporator at 55° C. before being resuspended in 100 μL 50% H_2O, 50% acetonitrile. This was analyzed with a Waters 2996 photodiode array detector Separations Module HPLC system, column used was C18-5 μM preparative column (4.6 × 250 mm, Waters, X-Bridge, Made in Ireland). The mobile phase was 0.1% acetic acid in water (A), and methanol (B) and a linear gradient was used; 95% A (5 min), 70% A (25 min), 0% A (30 min), 95% A (5 min), the flow rate was 1.0 mL/min. After identification through comparisons with authentic standards, based on retention time and UV spectrum, peaks were manually collected and the mass confirmed with mass spectroscopy.

β-Etherase Boosting Saccharification with Cellulase Enzymes

[0298] For saccharification reactions, biomass pretreated with β-etherase was incubated with 1.2 μg/mL enzyme cocktail (4:1 Celluclast: novo 188 (Novozymes)) in 50 mM sodium acetate at pH 4.5 and incubated overnight at 37-40° C. with shaking. This was performed alongside a control reaction with buffer only. Solids were removed by centrifugation, and residual protein was precipitated with 80% ethanol. The supernatant, containing mono- and oligosaccharides, was dried with a centrifugal evaporator before samples were resuspended in ultra-pure water and filtered through a 0.2 μm polytetrafluoroethylene (PTFE) filter. Five replicates from each sample were investigated, and carbohydrate composition was analyzed by high-performance anion-exchange chromatography (HPAEC).

High-Performance Anion-Exchange Chromatography (HPAEC)

[0299] High-performance anion-exchange chromatography was used to analyze monosaccharide release after saccharification. Briefly, 5 μL of samples or standards were injected on a CarboPac PA20 3 × 150 mm analytical column via a CarboPac PA20 3×0 mm guard column using Chromeleon 6.8 Chromatography Data Systems software (Dionex). Sugars were separated at a flow rate of 0.4-0.5 mL min^{-1} at a temperature of 25° C. as follows: after equilibration of the column with 100% H_2O, samples were separated in a linear gradient of 100% H_2O to 99%-1% of H_2O-0.2 M NaOH for 5 min, then constant for 10 min, followed by a linear gradi-

ent to 47.5%-22.5%-30% of H_2O-0.2 M NaOH-0.5 M NaOAc/0.1 M NaOH in 7 min and then kept constant for 15 min. After washing the column with 0.2 M NaOH for 8 min it was re-equilibrated with 100% H_2O for 10 min before the injection of the next sample. Carbohydrates were detected by ICS-3000 PAD system with an electrochemical gold electrode, identified by comparison with retention times of external standards (arabinose, fucose, galactose, glucose, glucuronic acid, mannose, rhamnose, and xylose) and quantified through the integration of these known standards.

Lignin Isolation

[0300] Enzyme lignins, representing essentially all of the lignin in the sample, were prepared following ball-milling of the cell wall isolate as previously described.(75-77, 78)

NMR Analysis

[0301] 2D NMR of enzyme lignins (EL) in 4:1 v/v DMSO-d_6:pyridine-d_5 were acquired on a Bruker Biospin (Billerica, MA) Avance 700 MHz spectrometer equipped with a **5**-*mm* QCI $^{1}H/^{31}P/^{13}$ C/^{15} N QCI cryoprobe with inverse geometry (proton coils closest to the sample), as described previously.(76,77) Volume-integration of contours in HSQC plots used TopSpin 4.07 (Mac version) software, and no correction factors were used. The data represent volume-integrals only, and data are presented on an S + G + H = 100% basis (FIG. **17**); pCA, and tricin T units are always terminal and are, therefore, likely overestimated.(77) Data assignments here were made by comparison with published data from other samples from our lab, including in the various tricin-related papers.(71-74, 79, 80)

Statistical Analysis

[0302] Where mentioned two tail ANOVAs were performed using R core package "stats".(83)

Example 1 Isolation of *Parascedosporium Putredinis* NO1

[0303] We inoculated liquid cultures containing wheat straw as the sole carbon source with samples of wheat straw-enriched compost and tracked the dynamics of the resulting microbial community using targeted amplicon sequencing during cultivation. Sequencing of 16S ribosomal RNA genes generated over three million reads from the prokaryotic community over the whole time course, which clustered together to form 25,304 operational taxonomic units (OTUs) (FIG. **1***a*). The most abundant bacterial phyla identified were the gram-negative Bacteroidetes, Verrucomicrobia and Proteobacteria, respectively, representing an average of 31%, 19.8%, and 15.5% of the total reads across the time course. Analysis of the eukaryotic community by sequencing the Internal Transcribed Spacer (ITS) region predominantly yielded reads that had no match within the UNITE fungal rDNA sequence database.23,24 In total, 96.5% of generated OTUs were not recognized as fungal and instead showed the closest homologies to protozoa. Among the fungi, we noted distinct changes in the composition of the community with time. In particular, a fungus (designated strain NO1) identified as *Parascedosporium putredinis* an Ascomycete in the Microascaceae family, showed increased abundance after 4 weeks of incubation (FIG. **1***b*). This fungus was readily isolated from shake flasks by culturing on both nutrient agar and potato dextrose agar and dominated the eukaryotic community in the shake flasks after four weeks of incubation, representing 84% of the identifiable fungal reads at 8 weeks, a time point by which, we hypothesize, the majority of easily accessible carbon from wheat straw has been depleted.[25] Interestingly, this fungus could be selectively cultivated when agar plates contained kraft lignin as the sole carbon source.

Example 2 Omics Analysis of Wheat Straw Degradation by *P. Putredinis* NO1

[0304] We confirmed that *P. putredinis* NO1 could grow on wheat straw as a sole carbon source and optimized the composition of growth media for cellulase and xylanase production using a central composite design (FIG. **6**). The deconstruction of wheat straw by *P. putredinis* NO1 over 28 days was subsequently tracked by measuring mass loss and carbohydrate-active enzyme (CAZy) activity (FIG. **7**). From this study, we identified the second, fourth and tenth day of incubation on wheat straw as distinct time points to harvest RNA for sequencing on an Illumina platform. These incubation times were chosen as together they represent the first detection of lignocellulolytic activity (day 2), the peak of enzyme activities (day 4) and the subsequent reduction of lignocellulolytic activity - a point at which the easily accessible sugars in the wheat straw had been utilized. In total, 5,586 unique contiguous DNA sequences (contigs) were assembled from the 339,854,704 reads generated, and differential gene analysis identified 2,189 contigs that were upregulated at high confidence and fold change ($P<0.001$, FC >10) when *P. putredinis* NO1 was grown on wheat straw compared to growth on glucose. These highly upregulated genes included those coding for 102 putative CAZy proteins; comprising 47 glycoside hydrolases (GH), 41 auxiliary activities (AA), ten carbohydrate esterases (CE) and a polysaccharide lyase (PL). The majority of CAZy family proteins were upregulated after four days of growth (FIG. **2**), in agreement with the peak of the observed enzymatic activities in *P. putredinis* NO1 culture supernatants.

[0305] As the macromolecular structure of lignocellulose prohibits intracellular degradation, many enzymes for its deconstruction must be secreted. We therefore performed LC-MS/MS analysis on protein samples collected directly from the culture supernatant, and separately from those bound to insoluble components of the culture using a biotin-labelling method designed to enrich for proteins tightly bound to the residual biomass.[26] We identified 3,671 proteins across all samples, including 1,037 proteins present in only wheat straw conditions (FIG. **8***a*). Within the resultant protein library, 275 sequences contained a recognizable CAZy domain. These accounted for 25.7% (194 proteins) of the molar percentage of the supernatant samples and 14.1% (174) of the biotin-labelled samples after four days of growth on wheat straw, compared to 13.3% (97) of the supernatant and 2% (56) of the biotin labelled samples from glucose-grown cultures (FIG. **8***b*).

[0306] The most abundant CAZy protein family, accounting for 3.7% and 3.6% of the respective supernatant and biotin-labelled fractions on the fourth day, were GH6s, which may be endoglucanases or processive cellobiohydrolases. These, along with GH7s, often constitute the major

cellulases in filamentous fungi.[27] The GH6 family, is represented by four distinct proteins within the proteome, included the most abundant single protein - c7229_g3_i1_1, a putative cellobiohydrolase with an 85.89% sequence identity to a cellulase (XP_016646396.1) from Scedosporium apiospermum. Other abundant GHs likely active on cellulose include GH7 (typically cellobiohydrolases or endoglucanases), GH5 and GH45 (often endoglucanases) and GH1 and 3 (typically glucosidases).[28]

[0307] Efficient lignocellulose deconstruction demands a combination of cellulolytic and hemicellulolytic enzymes that work cooperatively. Enzymes related to the depolymerization of arabinoxylan (major hemicellulose of wheat straw), were well represented within the exoproteome. Nine proteins were identified with homology to endo β-1-4-xylanases (GH10 and GH11), which hydrolyse the arabinoxylan backbone, and five proteins were identified as putative β-1,4-xylosidases that act on the resultant fragments to produce xylose monomers (GH3, GH31, GH43_1, GH43_11, GH43_36). Also of note were the GH43 subfamilies GH43_1, GH43_21, GH43_22, GH43_26 and GH43_36 that were abundant within the secretome, including putative β-D-xylosidases, α-L-arabinofuranosidase, and β-1,3-galactosidase activities. Fifteen GH43 subfamily members were identified, with nine proteins showing closest homology to known arabinofuranosidases.

[0308] Three proteins, belonging to the CE1 family, showed significant sequence homology to feruloyl esterases. Ferulic acid is esterified to the arabinose side chain of arabinoxylans, and through the formation of diferulate bridges and ester-ether linkages allows the respective formation of covalent interactions between arabinoxylan chains and lignin. Feruloyl esterases, therefore, are thought to aid the solubilization of plant cell wall polysaccharides by the hydrolysis of the ester link that exists between ferulic acid residues and arabinose, thereby disrupting the crosslinking of cell wall components.[29] Putative acetyl xylan esterases (3 in CAZy family CE1 and 3 in CE5) were also observed and are known to facilitate the degradation of xylan through the removal of acetyl substitutions.[30]

[0309] The CAZy auxiliary activity (AA) class is classified as containing enzymes that act in conjunction with carbohydrate-active enzymes through redox activities. Interestingly, 69 putative proteins from the AA class were detected in the exosecretome, more than many lignocellulose-degrading fungi contain in their total genome,[31] suggesting an important role for the oxidative degradation of lignocellulose in *P. putredinis* NO1. The AA9 family, which along with the AA10, AA11, AA13, AA14 and AA15 families constitute the lytic polysaccharide monooxygenases (LPMOs) - a class of copper metalloenzymes that catalyse the oxidative cleavage of glycosidic bonds in multiple polysaccharide substrates including chitin, cellulose, and xylan,[32.33] were highly represented within the exosecretome. In total, we identified nineteen putative LPMOs (16 AA9s; 2 AA11s; 1 AA13), fifteen of which were upregulated tenfold or more between glucose and wheat straw conditions. Fittingly, 16 AA3s (glucose-methanol-choline (GMC) oxidoreductase) and 9 AA7s (glucooligosaccharide oxidase), which have been shown to facilitate the activity of the LPMOs through electron shuttling,[34,35] were also present within wheat straw cultures.

[0310] Five putative multicopper oxidase proteins were also observed - two from the AA1_3 subfamily (Laccase-like multicopper oxidase) and one from the AA1_2 subfamily (Ferroxidase). Laccase-like multicopper oxidases are of unknown function but have been implicated in lignin degradation, as well as other diverse functions (iron homeostasis, offense/defence),[36] whereas ferroxidases have been reported to be involved in lignocellulose degradation in Ascomycetes, in which they generate hydroxyl radicals via the Fenton reaction.[37] Established lignin depolymerizing enzymes associated with the white-rot fungal decay of lignin, including laccases from the AA1_1 subfamily or peroxidases from the AA2 family, were not present within the libraries, perhaps not surprising given the *P. putredinis* NO1 sits within the Ascomycota phylum, and as such is closer in relation to the soft-rots.

[0311] Despite the apparent lack of known ligninases in *P. putredinis* NO1, a putative AA6 (1,4-benzoquinone reductase) associated with the intracellular biodegradation of aromatic compounds was present within the supernatant and may have a role in the metabolism of lignin breakdown products.[31,38]

[0312] Of key interest to us was the potential of *P. putredinis* NO1 to produce novel lignocellulolytic activities, particularly those able to boost lignocellulose deconstruction via the modification and solubilization of lignin. An unknown protein, c2092, identified in the exosecretome was subsequently found to have β-etherase activity and no CAZy identification.

Example 3 A New β-Etherase

[0313] The β-ether motif with its characteristic β—O—4 inter-unit linkage is the most abundant in lignin, estimated at representing over 50% of the total inter-unit linkages.[39] Enzymes employing β-ether cleavage mechanisms can deconstruct synthetic and extracted lignin;[40,41,42] these bacterial etherases that have been characterized to date, however, are intracellular proteins, and are glutathione- or NAD^+- dependent, suggesting that in nature they are not directly involved in the breakdown of the lignin macromolecule, but rather its smaller, membrane-transportable oligomers. An extracellular fungal protein displaying β-etherase activity was previously purified from the supernatant of the *Chaetomium* sp. 2BW- 1, although its identity remains unknown.[43]

[0314] Using a synthetic lignin model compound GGβ4MU (7-[2-hydroxy-2-(4-hydroxy-3-methoxyphenyl)-1-(hydroxymethyl)ethoxy]-4-methyl-2/7-1-benzopyran-2-one) containing a β-methylumbelliferyl ether, guaiacylglycerol-β-(4-methylumbelliferyl) ether (FIG. **9**),[44] that when cleaved yields the fluorogenic product 4-methylumbelliferone (4MU), we detected β-etherase activity within the culture supernatant of *P. putredinis* NO1. This activity was present when *P. putredinis* NO1 was grown on wheat straw but not on glucose, suggesting a possible role in lignocellulose degradation, and appeared to be independent of cofactors such as glutathione or NAD^+. Given its presence in the secretome and its apparent cofactor independence, we hypothesized that this putative ligninase was unlikely to share significant sequence homology to the previously described intracellular β-etherases from sphingomonads, and indeed no proteins with similarity to these enzymes were detected. We, therefore, subjected the culture supernatant of *P. putredinis* NO1 grown on wheat straw to a series of

protein fractionation techniques, enriching at each step for β-etherases activity.

[0315] The putative β-etherase was initially purified by ammonium sulfate precipitation of the proteins in the culture supernatant to decrease sample pigmentation and reduce protein-protein interactions. This treatment facilitated further purification by size-exclusion and anion-exchange chromatography. Using shotgun proteomics, we identified c2092, a 44.5 kDa protein present in the purified fraction that contained a predicted signal peptide. Analysis of the transcriptomic and proteomic data revealed this protein was strongly upregulated in the presence of wheat straw and present in both the supernatant and biotin-labelled proteomic libraries throughout the growth of *P. putredinis* NO1 on wheat straw (FIG. **10**). Using profile Hidden Markov models constructed by HMMER3 on using the pFAM database,[45] we saw homology to a common central tyrosinase domain (PF00264; Evalue = 7.1e-49) with a characteristic binuclear type-3 copper-binding site consisting of six histidine residues located in a four-helical bundle coordinating the binding of two copper ions[46] (FIG. **11**). Fungal tyrosinases are associated with pigmentation and browning; specifically, through melanin production, whereby they catalyse the introduction of a hydroxyl group at the ortho-position of a para-substituted monophenols and the subsequent oxidation to the corresponding o-quinone.[47] However, c2092 lacks both the C- and N-terminal domains that tyrosinases typically contain and instead shows higher homology (170/370 identity (46%)) to a catechol oxidase (AoCO4) from *Aspergillus oryzae*.[48] Catechol oxidases differ from tyrosinases due to a lack of mono-oxygenase activity.[49] Examination of the proteomics library resulted in the identification of seven sequences with significant similarities to c2092 (Table 1), all predicted to be extracellular and soluble, and five upregulated in the presence of wheat straw (FIG. **12**). Searches within the NCBI non-redundant database further revealed the presence of proteins of similar sequence (>50% sequence identity) distributed throughout fungal genomes of the Sordariomycetes class of Ascomycetes (Table 2).

Example 4 Experimental Confirmation of β-Etherase Activity

[0316] To determine if c2092 was responsible for the observed β-etherase activity, we heterologously expressed the codon-optimized sequence in *Escherichia coli*. The recombinant protein was purified (Table 3), and the β-etherase activity of the protein was confirmed by determining the level of fluorescence released after incubation with GGβ4MU (FIG. **13***a*). The pH and temperature dependency of the enzyme were investigated, revealing maximum activity at pH 10 and 60° C. (FIGS. **13***b-c*). Whereas the mushroom tyrosinase (*Agaricus bisporus*) has been reported to have promiscuous β-etherase activity on small synthetic compounds, no significant activity has been reported against macromolecular lignin.[50] The β-etherase from *P. putredinis* NO1 did not display activity against L-tyrosine and L-DOPA, as is characteristic of tyrosinases (FIG. **14**).[51] We subsequently assayed for potential oxidase activity against a range of phenolic substrates, including di-phenolics, known to be catechol oxidase substrates,[49] and observed no similarities to catechol oxidase in terms substrate preferences (FIG. **15**, Table 4). Interestingly, the etherase showed activity with the substrates: 4-hydroxybenzoic acid, vanillic acid, and quercetin, all known to be tyrosinase inhibitors.[52]

Example 5 Release of Tricin and Lignin Units from Wheat Straw

[0317] Tricin has recently been described as a subunit in the lignin of monocot species, incorporated through a 4—O—β linkage.[11] As wheat straw contains relatively high concentrations of tricin compared to other agriculturally relevant feedstocks,[8] we assessed the ability of the β-etherase to release tricin from wheat straw. The β-etherase was incubated with wheat straw for sixteen hours under physiological conditions (pH 8.5 and 30° C.). Reaction products were monitored by High-Performance Liquid-Chromatography (HPLC), and a peak corresponding to tricin was identified by reference to an authentic standard and confirmed by mass spectrometry. Under the growth conditions used for *P. putredinis* NO1, a significantly higher concentration of tricin was present in the reaction supernatant of wheat straw with the β-etherase compared to incubations with buffer alone (ANOVA, $F(2,12)=44.67$, $p<0.05$) (FIG. **4***a*). We were also able to detect the presence of p-coumaric acid, vanillin, and p-hydroxybenzaldehyde in the reaction supernatant through comparisons with authentic standards and mass spectrometry; however, unlike tricin, these compounds were not enriched under the β-etherase-treated reaction conditions (FIG. **16***c*) and presumably are produced as a result of simple ester cleavage.

[0318] NMR (FIG. **17**) of the enzyme lignins (EL) isolated (following crude polysaccharidase treatment to saccharify most of the polysaccharides),(75) and the product generated from it by a non-optimized treatment with our enzyme showed little change to the actual lignin profile but a strong decrease in the tricin level. Thus, even though integration of correlation contours in the spectra resulting from such 2D-HSQC (heteronuclear single-quantum coherence) experiments does not provide reliable quantification, their relative values are considered to be valid.(76,77) Analysis showed that the relative tricin ether level in the lignin dropped from nearly 12% in the control to about 8.5% after the treatment. We were initially disappointed that we couldn't detect similar reductions in levels of the β-ether units A (FIG. **17**), but caution that these are 'quantified' on an A+B+C=100% basis and it is easy to speculate on how the levels might not significantly change even with some (presumably low-level) β-ether cleavage. In spectra from the whole cell wall component (and not just the isolated lignin, not shown), the trends were similar and the T6 and T8 contours were particularly weak in the treated sample whereas the T2'/6' peak was relatively strong; we have noted this occurrence before in rapidly relaxing samples, and do not fully understand its origin; regardless, the relative tricin level in the treated material was again lower than in the control and obviously consistent with the measured release of tricin noted above.

[0319] We further tested the activity of the β-etherase on alternative feedstocks, including sugarcane bagasse and rice straw. A smaller amount of tricin was released from sugarcane bagasse compared to wheat straw; however, in contrast to assays with wheat straw, p-coumaric acid was significantly enriched (ANOVA, $F(2,12)=44.67$, $p<0.05$) (FIG. **4***b*, FIG. **16**). Rice straw showed little difference in product release, with relatively low concentrations of tricin and p-coumaric acid released during the incubation (FIG. **16**).

[0320] As mushroom tyrosinase has been reported to cleave β-ether linkages promiscuously,[50] we tested its β-etherase activity on these lignocellulosic substrates under equivalent conditions. We observed less tricin, p-coumaric acid, and p-hydroxybenzaldehyde production in the reaction mixtures containing mushroom tyrosinase compared to the *P. putredinis* NO1 β-etherase treatments. Tricin is a known tyrosinase inhibitor that binds non-competitively to the hydrophobic pocket of the protein,[53] and p-coumaric acid has been characterized as having a mixed-type inhibition effect.[54] This inhibition, through the non-reversible binding of the reaction products, could go some way to explaining why mushroom tyrosinase displays little activity towards the lignin macromolecule.

Example 6 β-Etherase Pretreatment Boosts Saccharification

[0321] The recalcitrance of lignocellulose to degradation requires that feedstocks are pretreated in order to disrupt lignin, before efficient saccharification can be achieved using current commercial enzymatic cocktails. These pretreatments are typically physico-chemical, using a combination of heat and pressure with acid, alkali or organic solvents. As these industrial processes are energy-intensive and environmentally damaging, the use of biological treatments, performed under relatively mild conditions, are desirable. To investigate if the application of the β-etherase would improve saccharification rates, we treated sugarcane bagasse, wheat straw, and rice straw with β-etherase for sixteen hours before the addition of commercial cellulases. Sugarcane bagasse demonstrated a major improvement in digestibility after pretreatment with β-etherase resulting in a significant increase in glucose, xylose, and arabinose compared to the untreated control (2-fold, 5-fold and 23-fold, respectively) after saccharification (FIGS. **5***a-b*). Wheat straw treated with β-etherase also showed an improvement in glucose release (ANOVA, $F(2,12)= 4.47$, $p<0.05$), albeit at a more modest level with a 1.2-fold increase. Interestingly, no improvement in saccharification was observed with rice straw, which may reflect the lower lignin content of rice straw compared to wheat straw and sugarcane.[55] This suggests that although the β-etherase can modify the plant cell wall structure and enhance digestibility, differences in lignocellulose organization and lignin content between feedstocks may determine the extent to which this occurs.

Example 7: Enzyme Homology and Identification

[0322] *P. putredinis* NO1 is able to dominate cultures in the latter stages of wheat straw degradation in a mixed microbial community, in liquid culture, when easily accessible polysaccharides have been exhausted. Using a combination of omics approaches, we have identified a diverse range of potentially industrially relevant carbohydrate-active enzymes, including a large number of enzymes associated with the oxidative attack on lignocellulose. In particular, we have identified a new extracellular β-etherase that is preferentially expressed in the presence of wheat straw and demonstrated that this enzyme can boost enzymatic hydrolysis by cellulases as well as selectively release the pharmaceutically relevant flavonoid tricin from monocot lignin. The cleavage of β-ether bonds most likely aids the breakdown of lignocellulose in natural environments. We contend that this ability to deconstruct and modify lignin is important for *P. putredinis* NO1 to be able to out-compete other microbial species during the latter stage of plant biomass degradation. Preferential removal of tricin subunits has been described by the white-rot fungi, Pleurotus eryngii, during the selective delignification of wheat straw and has been proposed to be key to lignocellulose degradation, although the enzyme activity that facilitated tricin release was not identified.[56] When the publicly available genome of *P. eryngii* was examined for the presence of proteins with homology to the β-etherase from *P. putredinis* NO1 no significant hits were detected. As the protein described as being responsible for β-etherase activity from *Chaetomium* sp. 2BW-1 was not identified to sequence level, it is unclear whether it shares homology to the enzyme described here; however, the proteins appear to be distinct as the reported sizes differ by 20 kDa.[43] Taken together, these observations suggest that multiple, structurally dissimilar, enzymes in the natural environment may mediate ether linkage disruption in lignocellulose-degrading microbes. To the best of our knowledge, this is the first identification and characterization of an extracellular β-etherase that has no cofactor requirement for activity capable of selectively releasing tricin from lignin and could have potential biotechnological applications.

REFERENCES

[0323] 8. Lan W, et al. Tricin-lignins: occurrence and quantitation of tricin in relation to phylogeny. 88, 1046-1057 (2016).

[0324] 11. Li M, pu Y, Yoo CG, RagauskasA. The occurrence of tricin and its derivatives in plants. *Green Chem* 18, (2016).

[0325] 23. Kõljalg U, et al. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. 166, 1063-1068 (2005).

[0326] 24. Abarenkov K, et al. The UNITE database for molecular identification of fungi - recent updates and future perspectives. 186, 281-285 (2010).

[0327] 25. Alessi AM, et al. Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. *Biotechnol Biofuels* 11, 166 (2018).

[0328] 26. Alessi AM, et al. Revealing the insoluble metasecretome of lignocellulose-degrading microbial communities. *Scientific reports* 7, 2356 (2017).

[0329] 27. Jun H, Guangye H, Daiwen C. Insights into enzyme secretion by filamentous fungi: Comparative proteome analysis of *Trichoderma reesei* grown on different carbon sources. J*ournal of Proteomics* 89, 191-201 (2013).

[0330] 28. Glass NL, Schmoll M, Cate JHD, Coradetti S. Plant cell wall deconstruction by ascomycete fungi. 67, 477-498 (2013).

[0331] 29. de Oliveira DM, et al. Ferulic acid: a key component in grass lignocellulose recalcitrance to hydrolysis. *Plant biotechnology journal* 13, 1224-1232 (2015).

[0332] 30. Zhang J, Siika-Aho M, Tenkanen M, Viikari L. The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed. *Biotechnol Biofuels* 4, 60 (2011).

[0333] 31. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire

of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* 6, 41 (2013).

[0334] 32. Vaaje-Kolstad G, et al. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* 330, 219-222 (2010).

[0335] 33. Quinlan RJ, et al. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. 108, 15079-15084 (2011).

[0336] 34. Laurent C, Breslmayr E, Tunega D, Ludwig R, Oostenbrink C. Interaction between cellobiose dehydrogenase and lytic polysaccharide monooxygenase. *Biochemistry* 58, 1226-1235 (2019).

[0337] 35. Tan T-C, et al. Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. *Nature Communications* 6, 7542 (2015).

[0338] 36. Levasseur A, et al. Exploring laccase-like multicopper oxidase genes from the ascomycete Trichoderma reesei: a functional, phylogenetic and evolutionary study. *BMC Biochemistry* 11, (2010).

[0339] 37. Kersten P, Cullen D. Extracellular oxidative systems of the lignin-degrading basidiomycete *Phanerochaete chrysosporium*. *Fungal Genetics and Biology* 44, 77-87 (2007).

[0340] 38. Daly P, et al. Expression of *Aspergillus niger* CAZymes is determined by compositional changes in wheat straw generated by hydrothermal or ionic liquid pretreatments. *Biotechnol Biofuels* 10, 35 (2017).

[0341] 39. Schutyser W, Renders T, Van den Bosch S, Koelewijn SF, Beckham GT, Sels BF. Chemicals from lignin: an interplay of lignocellulose fractionation, depolymerisation, and upgrading. *Chemical Society reviews* 47, 852-908 (2018).

[0342] 40. Gall DL, et al. In vitro enzymatic depolymerization of lignin with release of syringyl, guaiacyl, and tricin units. *Applied and environmental microbiology* 84, (2018).

[0343] 41. Kontur WS, et al. A heterodimeric glutathione S-transferase that stereospecifically breaks lignin's β(R)-aryl ether bond reveals the diversity of bacterial β-etherases. *The Journal of biological chemistry* 294, 1877-1890 (2019).

[0344] 42. Marinovic M, et al. Selective cleavage of lignin β—O—4 aryl ether bond by β-etherase of the white-rot fungus *Dichomitus squalens*. *ACS Sustain Chem Eng* 6, 2878-2882 (2018).

[0345] 43. Otsuka Y, Sonoki T, Ikeda S, Kajita S, Nakamura M, Katayama Y. Detection and characterization of a novel extracellular fungal enzyme that catalyzes the specific and hydrolytic cleavage of lignin guaiacylglycerol β-aryl ether linkages. 270, 2353-2362 (2003).

[0346] 44. Weinstein DAG, M.H. Synthesis of guaiacylglycol and glycerol-β—O—(β-methylumbelliferyl) ethers: lignin model substrates for the possible fluorometric assay of β-etherases. *Holzforschung* 33, 134-135 (1979).

[0347] 45. Finn RD, et al. The Pfam protein families database. *Nucleic Acids Research* 38, D211-D222 (2010).

[0348] 46. Kanteev M, Goldfeder M, Fishman A. Structure-function correlations in tyrosinases. *Protein Science* 24, 1360-1369 (2015).

[0349] 47. Halaouli S, Asther M, Sigoillot JC, Hamdi M, Lomascolo A. Fungal tyrosinases: new prospects in molecular characteristics, bioengineering and biotechnological applications. *Journal of Applied Microbiology* 100, 219-232 (2006).

[0350] 48. Hakulinen N, Gasparetti C, Kaljunen H, Kruus K, Rouvinen J. The crystal structure of an extracellular catechol oxidase from the ascomycete fungus Aspergillus oryzae. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry* 18, 917-929 (2013).

[0351] 49. Gasparetti C, Faccio G, Arvas M, Buchert J, Saloheimo M, Kruus K. Discovery of a new tyrosinase-like enzyme family lacking a C-terminally processed domain: production and characterization of an *Aspergillus oryzae* catechol oxidase. *Applied Microbiology and Biotechnology* 86, 213-226 (2010).

[0352] 50. Min K, et al. Perspectives for biocatalytic lignin utilization: cleaving 4-O-5 and Cα—Cβ bonds in dimeric lignin model compounds catalyzed by a promiscuous activity of tyrosinase. *Biotechnol Biofuels* 10, 212 (2017).

[0353] 51. Yang Z, Robb DA. Comparison of tyrosinase activity and stability in aqueous and nearly nonaqueous environments. *Enzyme and Microbial Technology* 15, 1030-1036 (1993).

[0354] 52. Zolghadri S, et al. A comprehensive review on tyrosinase inhibitors. *J Enzyme Inhib Med Chem* 34, 279-309 (2019).

[0355] 53. Mu Y, Li L, Hu S-Q. Molecular inhibitory mechanism of tricin on tyrosinase. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 107, 235-240 (2013).

[0356] 54. Lim JY, Ishiguro K, Kubo I. Tyrosinase inhibitory p-coumaric acid from ginseng leaves. *PhytotherRes* 13, 371-375 (1999).

[0357] 55. Van Soest PJ. Rice straw, the role of silica and treatments to improve quality. *Animal Feed Science and Technology* 130, 137-171 (2006).

[0358] 56. van Erven G, Nayan N, Sonnenberg ASM, Hendriks WH, Cone JW, Kabel MA. Mechanistic insight in the selective delignification of wheat straw by three white-rot fungal species through quantitative ^{13}C-IS py-GC-MS and whole cell wall HSQC NMR. *Biotechnol Biofuels* 11, 262 (2018).

[0359] 57. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods* 7, 335-336 (2010).

[0360] 58. DeSantis TZ, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology* 72, 5069-5072 (2006).

[0361] 59. McDonald D, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* 6, 610-618 (2012).

[0362] 60. Abarenkov K, et al. The UNITE database for molecular identification of fungi-recent updates and future perspectives. *The New phytologist* 186, 281-285 (2010).

[0363] 61. Bezerra MA, Santelli RE, Oliveira EP, Villar LS, Escaleira LA. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta* 76, 965-977 (2008).

[0364] 62. Lever M. Colorimetric and fluorometric carbohydrate determination with p-hydroxybenzoic acid hydrazide. *Biochemical Medicine* 7, 274-281 (1973).

[0365] 63. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* 39, W29-W37 (2011).

[0366] 64. Camacho C, et al. BLAST+: architecture and applications. *BMC bioinformatics* 10, 421 (2009).

[0367] 65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology* 215, 403-410 (1990).

[0368] 66. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature protocols* 2, 953-971 (2007).

[0369] 67. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* 8, 785-786 (2011).

[0370] 68. Ishihama Y, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & cellular proteomics : MCP*4, 1265-1272 (2005).

[0371] 69. H. Pages PA, R. Gentleman and S. DebRoy. BioStrings: Efficient manipulation of biological strings. (ed^(eds). R package version 2.52.0. edn (2018).

[0372] 70. Karambelkar P, Jadhav, V.M. , Kadam, V. Isolation and characterization of flavonoid tricin from sugarcane sludge. *Indo American Journal of Pharmaceutical Research* 4, 7 (2014).

[0373] 71. J. C. del Rio et al., Structural characterization of wheat straw lignin as revealed by analytical pyrolysis, 2D-NMR, and reductive cleavage methods. Journal of Agricultural and Food Chemistry 60, 5922-5935 (2012).

[0374] 72. W. Lan et al., Tricin-lignins: Occurrence and quantitation of tricin in relation to phylogeny. Plant J. 88, 1046-1057 (2016).

[0375] 73. W. Lan et al., Tricin, a flavonoid monomer in monocot lignification. Plant Physiol. 167, 1284-U1265 (2015).

[0376] 74. W. Lan et al., Maize tricin-oligolignol metabolites and their implications for monocot lignification. Plant Physiol. 171, 810-820 (2016).

[0377] 75. H.-M. Chang, E. B. Cowling, W. Brown, E. Adler, G. Miksche, Comparative studies on cellulolytic enzyme lignin and milled wood lignin of sweetgum and spruce. Holzforschung 29, 153-159 (1975).

[0378] 76. H. Kim, J. Ralph, Solution-state 2D NMR of ball-milled plant cell wall gels in DMSO-d_6/pyridine-d_5. Org Biomol Chem 8, 576-591 (2010).

[0379] 77. S. D. Mansfield, H. Kim, F. Lu, J. Ralph, Whole plant cell wall characterization using solution-state 2D-NMR. Nature protocols 7, 1579-1589 (2012).

[0380] 78. H. Kim et al., Monolignol benzoates incorporate into the lignin of transgenic *Populus trichocarpa* depleted in C3H and C4H. ACS Sustain Chem Eng 8, 3644-3654 (2020).

[0381] 79. J. Rencoret et al., Structural characterization of lignin isolated from coconut (*Cocos nucifera*) coir fibers. Journal of Agricultural and Food Chemistry 61, 2434-2445 (2013).

[0382] 80. W. Lan et al., Elucidating tricin-lignin structures: Assigning correlations in HSQC spectra of monocot lignins. Polymers (Basel) 10, 916 (2018).

[0383] 81. H. Zhang et al., dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* 46, W95-W101 (2018).

[0384] 82. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* 39, W29-W37 (2011).

[0385] 83. R. C. Team (2019) R: A Language and Environment for Statistical Computing. (Vienna, Austria).

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 47

<210> SEQ ID NO 1
<211> LENGTH: 1227
<212> TYPE: DNA
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 1

atgccttctg cgaagcgtct tctcgggctc ctgctcgccg ccaccgcggc ggtcggtgta 60

gctgcccagg aacctgccct caccgaggat gatttcagca tccccgagat cgagggaggt 120

gatgccctcg cacagcttgc ccagctcgcg gcggattctt ctcaggagac tgccctcagg 180

atggcgaagc gcggtctgaa cagcggctgc agccctagcc agatcaaggt tagaagggaa 240

tggcgaaccc tgacttctgc tcagcgcaag cagtacattg cgtcggtcaa gtgtcttcag 300

acgaagccca gcttcttcga ccccaacatc atcccggccg ctaagtcgct atttgacgac 360

tttgttggtg tccatgtttt ccagactggg tccatccatc tcactgccac tttcctcaca 420

tggcatcgct acttcgtcta tacctacgag acgaagctcc gcgaggaatg cggttacact 480

-continued

```
ggacccttgc cctactggga gtggggacta gacgtcaaca accccaacgc ctccccggtc      540

ttcgacggct ccgacacgtc tctgagcggc aacggtgcct tctttgcgca tgagggcatc      600

cagatggtgc agcctatcaa cggcaacatc ctcaagctcc cccccggcaa cggtggtggc      660

tgcgtgacca agggtccctt caaggacatg aaggttcact ttggcaccat catcctgccc      720

gtgtacggcc agcctatcct cagtggtgtc gagaacccca ttgccgacaa cgagcgctgc      780

ctcaagcgcg atctcaatgc cggcatcgct aagcgcttca ctagcttcct caactcgacc      840

agcgtgattc tcaagaacaa caacatcgag atgttccagg cccatctgca gggcgacgac      900

cgctacgtgc tcaaccagct cggtgttcac ggtggtggtc actacaccat cggcggtgac      960

cccggtggtg atcccttcat ctcccctggt gacccggctt tctacctcca ccacgcccag     1020

attgaccgca tctactggat ctggcagatg ctcgacttca agaaccgtca gggtgtccac     1080

ggtaccgcca ccctccagaa caaccctccc agcgccaacg ttaccgttga ggacaccatc     1140

gacctctctc ccctcgcccc gcctgtcaag atcaaggacc tcatgaacac tgtcggaggc     1200

tcgcccttgt gctacatcta cctctaa                                        1227
```

<210> SEQ ID NO 2
<211> LENGTH: 1158
<212> TYPE: DNA
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 2

```
atgtccgtcg tcaagaagct cctcgccgcc ctcgcggcca ccactttcct caccggcgtc       60

gctgctcaga cctatgagtt tagcgaggag gagctcacct ctggtgatgc cctcaaggcc      120

cttagcaagc aggctatgga aaatgctctt gcccgcctcc ccgaatctgg agagggctgc      180

acgcgggaga acgtcaaaat tcgaaaggaa tggcgcaaca tgccggccga gatgagaatc      240

ggctatgtca gcgctctcca gtgcttgatg gaatccgaaa gcgaatatcc cgacgtcgac      300

ggtgccaaga cggcgtttga cgacttcgcc gttcttcatt acaacctcac gccgttcgtg      360

cataactctg ctaccttcct taccttccac aggtactaca ttcacaccct ggaagagcag      420

atgaggaaca agtgcggata cactggtgac ttcccctact gggagtgggg cctcgactgc      480

gacgacccgc aacagtctcc cctctttgac ggctccgaaa cctctctcgg cagcgacggt      540

gagcctgtgg aggccggagc cggcggtggc ttcggcggcg gcttcggctt tggcatggga      600

ggcggcagcg gtggtggctg cgtgatgaag ggacccttct ccaactacac cgtcaacctc      660

ggaccctcaa ccaccgccga cccgctcgca tataacccgc gctgcatcaa gcgaaacctg      720

aacggcgcta tctgcaagca gaatgcctcg ctccggaaca cgacgacgac gatccttgac      780

tcgcccgata tcgaactatt ccaggcgatt gtccaaggtg acatgcgata ccccgaggcg      840

aggggtcttg gcatggccgt ccacggcggt ggtcacttta ctattggtgg tgatcccggt      900

ggcgacttct acttctcccc tctcgagccc gccttcttcc aacaccacgg ccagatcgac      960

cgcatgtact ttgtctggca gaacctcgac tgggaaaccc ggcagaacat tgccggcacc     1020

ggtaccatga tgaaccagcc ccccagccca gaggttgaga tcaccgaact cctcgacctc     1080
```

-continued

agccccctttg cagaagccag gccgatcaag gacctcattg acacccttgg ctcggccccg 1140

ttctgctttg tttacgag 1158

<210> SEQ ID NO 3
<211> LENGTH: 1230
<212> TYPE: DNA
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 3

atgctcgtgt acgcttcgct ggcgatcctg ccccttcttg ccggggtggg cgcatcgccg 60

cttaacaaga aggccacctt ttcgtaccag caggtatctc gaaatcctga cttcccgttg 120

gatgtggtcg acgagctcga ggccaaagcc atgccggggg ttgagacatg gatggctggg 180

aaggacaacg ccaacggctg cacgcttgaa aatgctgcgg tccgtagaga atggggtgac 240

ttatccgttg ccgagcgtga agagtacgtt gccgcggttc tctgtctcca gaagttacca 300

tccaaggcgc ccgagggaaa ggcaccaggt gcactcagcc gcttcgatga cttcgttgcc 360

acgcacatga cacaggccat gatgcttcac tcgcccacca atctgtttgc gagccaccgt 420

tactatatct gggcctacga aaccgctctt cgtgaagagt gcggctatac gggctaccag 480

ccgtacatga attacgaccg ttatgccgat gacctcctca actccccctt gttcaacggg 540

aacgcgtcca gcctaggagg caacggagct ccaagccaat acgccggtgt ccctcagcca 600

ttcaggcctc cgtacaacat gatcccctct gctggaggcg gtggttgtgt gaccgagggt 660

ccattcaaag acatggtcgt gagtcttggc cccgtgggaa tcatcgtcaa tgatatcccc 720

aagaacccgc gctccgatgg cctcggctcc aacccgcgat gcctccgtcg ggacgtgaat 780

aaatttttcgg cggccggcgc caaggccaac tacacgtaca acctgattac ggagaacgcc 840

ggaattgatg cgttttataa ccgctacctc gggcagccgc agctgaagga tgatcccaat 900

ccttgggggc tccataatgc cggccactat atcgtaggcg gcgacccagg cggtgatttc 960

tacgcttccc ccggcgaccc ctacttctac ttccaccacg gcatgttgga ccgcgtgtgg 1020

tggatctggc agatgcagga cccggccgcc cgcgtcaacc ttattccggg taccggcgcc 1080

ccggcgatga accatcccgg catgccgatg aaccgtcggc agtcgtccgc gacaattgtg 1140

gacttgggtt ggacggcgcc ggcggtgccc attacggagc tcaacgattc cttgggcggt 1200

aacggcggca agttttgcta cgtgtacgtg 1230

<210> SEQ ID NO 4
<211> LENGTH: 1152
<212> TYPE: DNA
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 4

atgcgtctgt tcaaggctct cgccgccgcg gccctgaccg gcctcgtcgc ggccgatgct 60

gtcaatgacc tagagaccaa gggccgtgcc gccctcgacg ccgtcattga gagctctacc 120

acgtgtagca aggacaagct caaggtccgc agagagtggg gagatatcag caccaccgag 180

cgcaaggctt acctcgacgg agtgctgtgc ctcctgaaca caccctccaa gctcgatccc 240

-continued

gctcgttacc ccggcgccaa gaaccgctac gatgactttg tcgttgttca catgaaccag 300

accctctcca tccatggaac cggtaacttc ctcgtgtggc accgctacta cgtgtgggcc 360

tgggagaacg tcatgaggac cgagtgtggc tatgaaggaa cccagcccta ctgggactac 420

ggccgctggg ccgaggatcc tctttcctcg cctctcttcg acggaagcga gacttcgctt 480

ggcggaaacg gcgcccccgt aacccagaac aagcgcagcc gcgtggaagg ccgccagttc 540

ggcggtggcg gtggtttcgg tggtggcctt ggcggaggtt tcggaggcgg cggtgatggc 600

ggtggctgca tttccaccgg tcccttcaag gacatggtcg tcaccctcgg ccccatgtcc 660

gccgtcgtca ggcccgcacc ggcccgcaac ccccaggccg acggctacgg tagtaacccc 720

cgctgcatcc gccgcgacat caccaactcg ctgagcatgg cctacggaaa gaccgaggac 780

atcgtcaaca gcatcgtcaa ctacaacgac atccttgcct tccagaactt catgcagggc 840

ggtaccggcg tgcacggcgt cggccacttc accgtctccg gcgaccccgg tggtgatttc 900

tacatctccc ccaacgagcc ttccttctgg ctccaccacg ccatgattga ccgcatctgg 960

accatctggc agtcccagga ctacgagacc cgcaggggcg ccatggaggg aggcaccagc 1020

atgatgggag gtggcagggc ccagtccctc gatgacctcg tcgacctcgg cgtcattgcc 1080

gacactgtct accccatccg cgacatcctc agctctgttg acggccccgg ccccttctgc 1140

tacgtgtacg ag 1152

<210> SEQ ID NO 5
<211> LENGTH: 1185
<212> TYPE: DNA
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 5

atgcgttctg ctttggctct cgtactcgcg gcctctctcc tcggcggaga ggccagcagc 60

atcaagaagc gattctcgac actcgatgtt tggcgccacg gcgactacga gcgggatatc 120

gtcgatcagc tctccgacga aacgtttccc aagatcgccg agtgggtcga gaagaccggc 180

tcgacctgca ctctcgaaaa tgccgtgcag cgaaaggagt ggaccgattt gaccattgat 240

gagagggcgg actacatcca agcggtgcag tgcttgatga agctccctcc caagtcgcag 300

gaccaagttc ccggctccct caaccgatat gacgacttcg tggccactca cgttactggt 360

attccggttc ttcacgcacc taccaacctc ttcgccagcc acaggtacta tatctgggcg 420

tacgagttgg cacttcgcga ggagtgcggg tacaagggct accagccgta catgaactac 480

gagcgacacc aggatcccat cacctcgccc ctgttcaacg gaaatgccac cagcatgggc 540

ggcaatggag cggcggctga gtaccccggc gtagtcatgc cttatcccag gccctacaac 600

gtcattcccg ctgcaggcgg tggtggctgc gtcacggaag gtcccttctc cgacatggtt 660

gtcagcatcg gccctctggg cactgttctc cgcgacattc cccgcaaccc ccgcgccgat 720

ggcctcggtt ccaacccccg ctgcctgcga cgcgatctta acaagttctc cgccgctgga 780

gcctccgcca accactcgta ctcgctcatt atggactacc ccgatgtcga cgccttctac 840

aaccgatacc tcggacagcc gttcctgaga ggagatgaat tcccgtgggg tcttcactct 900

-continued

gccggtcact acatcacggg aggagaccca ggtggcgact tttacgcctc gcctggtgac 960

ccgaccttct ggatgcatca tgccgccctc gaccgcttgt ggtggctgtg gcagatgcag 1020

gatcccgaga accgcctgca ggccatcccc ggcatcactt cgtcgaggat gaccaacgag 1080

gatgctcaaa agacaatggt ggatctgaag tggacggcag agccccgctc gctcggagat 1140

ctcaacgatc aaatgggaag tgccccttc tgttatatct atgta 1185

<210> SEQ ID NO 6
<211> LENGTH: 1059
<212> TYPE: DNA
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 6

atgcagctca ccattctcgc aacggcgctt cttgccgtca gcgcctccgc agcgcccacc 60

tgctgcacca acgccgagcg gcgagagtgg agaactttct caactaagga gaagcaagcc 120

tacatcgcgg cagtgaaatg tcttcaaagc aagccatctc agttgaagag cacctacccg 180

acgtcgcaga atcgattcga tgactttcag gcggttcata ttgacctgac ggagaagtat 240

cactttactg gacctttcca ggcctggcat cgcgtctttc tccacaagta cgaatccgac 300

ctccggggac tttgtgcata caagggctac cagccatact gggactggac caaggattcc 360

gggtctgagg ctgccttcct cgcctcgcct gtctttgacg ctgttaatgg cttcggtgga 420

aacggacctt acgttgatac gtcgaacttt cccgtcacca acgtccccgt caaaatcccg 480

aataagactg gcggaggctg tgttcaagat ggcgcctttg tcaatatgac agtcactctc 540

ggcccaggcc ctagtcttga gtcgaacccc cgctgtctca cccgtgactt cagctactgg 600

ctcatctcac ggaccctaac caaggcggtg gtagactgga ctcttgaggc cgcgtctttc 660

gctgtctttg acttccgtct gcagggcaca ggcatcgagc cagaaggcat gacagttcat 720

gcaggtggcc accttggtgt tggtggtgat atcggcgaga tcggcaacat gtattcatcc 780

cccggagacc cgctcttcta tcttcaccac gccaatcttg acagactctg ggaccagtgg 840

cagaggaaga agtttgctca gcgtgtccag gatatgaccg gaccggatac aatgtgggcc 900

tatccattca acttctttgg cgacgtaccc tataccaata tcactctgga aaccttgctc 960

gacttcaagg gtctccttgg gtccagctcg gcagaccgat atgtcaagat caaggacgtt 1020

atggacagcc agggtccgaa cctctgcgtc ttctacaaa 1059

<210> SEQ ID NO 7
<211> LENGTH: 1167
<212> TYPE: DNA
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 7

atggtggcga tcagctacgt ccttacggcg ttggcggtgg cgatccctgc ccttgcgcag 60

ggtccgtgtt ccacaattcg tcaacgacgg gcatggcata cgctgagcaa cagcgagaaa 120

cgcgcttact tggatgccga ggtttgcttg ctgggcaagg cccccaagtt tggcttcgag 180

ggggccaaaa acaggttcga ggagctccag gccgctcatc aagtccaggc ttatctcatt 240

-continued

catggagttg gagcgttttt gcctttccac agatatctta tgcacgctca tgagaccctg 300

ttgaggactg agtgtgggta ccagggagcg caaccatact gggatgagac acgggatgcc 360

ggtcgcgtca gtgagtccga gatccttgat ccggacactg gtttcggagg agatggtgtg 420

ggtgagaggg gttgtatcgc cgacggaccg tttgcaggct acatcaacag catcgggcca 480

ggttatagga ttacggatcg ctgcatcaca cgcttcgtga acaacacccg aagcttgatg 540

gctagccccc gtttcaccga caggtgccaa ggcatgaatc gatatgttga cgtgtggcca 600

tgcctcgagg gaaaccctca taattcgggc catggcgcta ttagcggact gatgatggac 660

ccgatcgcca gccccggcga ccccatcttc taccttcatc acacatggct cgataagctg 720

tggtgggagt ggcaggccat ggacctcccc cgacgcctca ccgacattgg cggccgtaac 780

acccaggacg gctccgaagg tttccccggt gcaccgccca actctaacgg tcccaacggc 840

gcgaagcgac gatcccccgc cgacggcccc atcctcatct tccccgggga cggaggaaac 900

attgactggg acgagattga ctgggaaaag attggcttcc ccggcctcgg cggtggaaat 960

ggaggcccga tccagttgcc gcccggcgtt gacgttcccc cggaagccct ggaacctccc 1020

gaggatgccg agcctcaaga gcccaggggt gacccgggcg atgtgaccac gttgaaccac 1080

gtcctgaaga tgtttggcct cgttccagac gccttgatcc gagacgtcat ggatatcgcc 1140

ggaggcactc tgtgttacga gtacgtc 1167

<210> SEQ ID NO 8
<211> LENGTH: 1044
<212> TYPE: DNA
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 8

atgaaaaacc tcgcgggact aataacggcc ttggccaccg gcgttgggat ggcgcatacg 60

catacacttg tgccgcacat gcaagactca actccctgta tcaacccaag tttgcgacgt 120

ccatggcaga ttctctcgga cggcgaaaaa cgctcatatc tcgatgccca gctttgtgtg 180

atgagaacgc cgcagaccct cggtcttcct ggcgcgagaa cacgtttcga ggagctggct 240

gccacccacc agattggcgc ccgtgccagc catgccacgg ggacattttt cccctaccat 300

cgatacctac tgcatgccca tgagtcattg ctgaaggagt gcggctacca cgcaggtctc 360

ccttactggg atgagaccag ggaagctgga aatttcatca agtctaccat attcgaatcg 420

ggcctcggat tcggtggctt tggaagcgac ctcaaagggt gcatcgaaga cggacctttc 480

gcaaacttga caagtacaat cggtcccggg ttttcgctga acgaacactg catctcacgt 540

gcgctcaatg aaactgcagg gctcaaggcg gctagggaag aggttgataa gtgcttagaa 600

gccaacgact atacagagat gtggcgttgt gcatatacca caccccatcg tgggggtcat 660

gggggcgtgg gaggcacgat gggagacgct ttggcatcgc ccggcgaccc ggtattctac 720

gtccaccacg cttgggtcga taagatttgg tgggattggc aggaggctga tcttgataat 780

agaatgtatg ctattggcgg gcccagcttc cagtcacctg atatcgggtt tcctgaggtt 840

cctggtgatg ttgaggaaga agaagcaaat atctttggca aaccaagcga agcaatccga 900

-continued

cgactacagg agctatggag ttcctctgat ccaagcagag aaactaccct ggagcataat 960

ctgacattgc taggtattat tcccgacatc aacattagta aggtcatgga cactagaggt 1020

gggtatcttt gctacgaata tgtt 1044

<210> SEQ ID NO 9
<211> LENGTH: 408
<212> TYPE: PRT
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 9

Met Pro Ser Ala Lys Arg Leu Leu Gly Leu Leu Leu Ala Ala Thr Ala
1 5 10 15

Ala Val Gly Val Ala Ala Gln Glu Pro Ala Leu Thr Glu Asp Asp Phe
20 25 30

Ser Ile Pro Glu Ile Glu Gly Gly Asp Ala Leu Ala Gln Leu Ala Gln
35 40 45

Leu Ala Ala Asp Ser Ser Gln Glu Thr Ala Leu Arg Met Ala Lys Arg
50 55 60

Gly Leu Asn Ser Gly Cys Ser Pro Ser Gln Ile Lys Val Arg Arg Glu
65 70 75 80

Trp Arg Thr Leu Thr Ser Ala Gln Arg Lys Gln Tyr Ile Ala Ser Val
85 90 95

Lys Cys Leu Gln Thr Lys Pro Ser Phe Phe Asp Pro Asn Ile Ile Pro
100 105 110

Ala Ala Lys Ser Leu Phe Asp Asp Phe Val Gly Val His Val Phe Gln
115 120 125

Thr Gly Ser Ile His Leu Thr Ala Thr Phe Leu Thr Trp His Arg Tyr
130 135 140

Phe Val Tyr Thr Tyr Glu Thr Lys Leu Arg Glu Glu Cys Gly Tyr Thr
145 150 155 160

Gly Pro Leu Pro Tyr Trp Glu Trp Gly Leu Asp Val Asn Asn Pro Asn
165 170 175

Ala Ser Pro Val Phe Asp Gly Ser Asp Thr Ser Leu Ser Gly Asn Gly
180 185 190

Ala Phe Phe Ala His Glu Gly Ile Gln Met Val Gln Pro Ile Asn Gly
195 200 205

Asn Ile Leu Lys Leu Pro Pro Gly Asn Gly Gly Gly Cys Val Thr Lys
210 215 220

Gly Pro Phe Lys Asp Met Lys Val His Phe Gly Thr Ile Ile Leu Pro
225 230 235 240

Val Tyr Gly Gln Pro Ile Leu Ser Gly Val Glu Asn Pro Ile Ala Asp
245 250 255

Asn Glu Arg Cys Leu Lys Arg Asp Leu Asn Ala Gly Ile Ala Lys Arg
260 265 270

Phe Thr Ser Phe Leu Asn Ser Thr Ser Val Ile Leu Lys Asn Asn Asn
275 280 285

-continued

```
Ile Glu Met Phe Gln Ala His Leu Gln Gly Asp Asp Arg Tyr Val Leu
    290                 295                 300

Asn Gln Leu Gly Val His Gly Gly Gly His Tyr Thr Ile Gly Gly Asp
305                 310                 315                 320

Pro Gly Gly Asp Pro Phe Ile Ser Pro Gly Asp Pro Ala Phe Tyr Leu
                325                 330                 335

His His Ala Gln Ile Asp Arg Ile Tyr Trp Ile Trp Gln Met Leu Asp
            340                 345                 350

Phe Lys Asn Arg Gln Gly Val His Gly Thr Ala Thr Leu Gln Asn Asn
        355                 360                 365

Pro Pro Ser Ala Asn Val Thr Val Glu Asp Thr Ile Asp Leu Ser Pro
    370                 375                 380

Leu Ala Pro Pro Val Lys Ile Lys Asp Leu Met Asn Thr Val Gly Gly
385                 390                 395                 400

Ser Pro Leu Cys Tyr Ile Tyr Leu
                405


<210> SEQ ID NO 10
<211> LENGTH: 386
<212> TYPE: PRT
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 10

Met Ser Val Val Lys Lys Leu Leu Ala Ala Leu Ala Ala Thr Thr Phe
1               5                   10                  15

Leu Thr Gly Val Ala Ala Gln Thr Tyr Glu Phe Ser Glu Glu Glu Leu
            20                  25                  30

Thr Ser Gly Asp Ala Leu Lys Ala Leu Ser Lys Gln Ala Met Glu Asn
        35                  40                  45

Ala Leu Ala Arg Leu Pro Glu Ser Gly Glu Gly Cys Thr Arg Glu Asn
    50                  55                  60

Val Lys Ile Arg Lys Glu Trp Arg Asn Met Pro Ala Glu Met Arg Ile
65                  70                  75                  80

Gly Tyr Val Ser Ala Leu Gln Cys Leu Met Glu Ser Glu Ser Glu Tyr
                85                  90                  95

Pro Asp Val Asp Gly Ala Lys Thr Ala Phe Asp Asp Phe Ala Val Leu
            100                 105                 110

His Tyr Asn Leu Thr Pro Phe Val His Asn Ser Ala Thr Phe Leu Thr
        115                 120                 125

Phe His Arg Tyr Tyr Ile His Thr Leu Glu Glu Gln Met Arg Asn Lys
    130                 135                 140

Cys Gly Tyr Thr Gly Asp Phe Pro Tyr Trp Glu Trp Gly Leu Asp Cys
145                 150                 155                 160

Asp Asp Pro Gln Gln Ser Pro Leu Phe Asp Gly Ser Glu Thr Ser Leu
                165                 170                 175

Gly Ser Asp Gly Glu Pro Val Glu Ala Gly Ala Gly Gly Gly Phe Gly
            180                 185                 190
```

-continued

```
Gly Gly Phe Gly Phe Gly Met Gly Gly Gly Ser Gly Gly Gly Cys Val
        195                 200                 205

Met Lys Gly Pro Phe Ser Asn Tyr Thr Val Asn Leu Gly Pro Ser Thr
    210                 215                 220

Thr Ala Asp Pro Leu Ala Tyr Asn Pro Arg Cys Ile Lys Arg Asn Leu
225                 230                 235                 240

Asn Gly Ala Ile Cys Lys Gln Asn Ala Ser Leu Arg Asn Thr Thr Thr
                245                 250                 255

Thr Ile Leu Asp Ser Pro Asp Ile Glu Leu Phe Gln Ala Ile Val Gln
            260                 265                 270

Gly Asp Met Arg Tyr Pro Glu Ala Arg Gly Leu Gly Met Ala Val His
        275                 280                 285

Gly Gly Gly His Phe Thr Ile Gly Gly Asp Pro Gly Gly Asp Phe Tyr
    290                 295                 300

Phe Ser Pro Leu Glu Pro Ala Phe Phe Gln His His Gly Gln Ile Asp
305                 310                 315                 320

Arg Met Tyr Phe Val Trp Gln Asn Leu Asp Trp Glu Thr Arg Gln Asn
                325                 330                 335

Ile Ala Gly Thr Gly Thr Met Met Asn Gln Pro Pro Ser Pro Glu Val
            340                 345                 350

Glu Ile Thr Glu Leu Leu Asp Leu Ser Pro Leu Ala Glu Ala Arg Pro
        355                 360                 365

Ile Lys Asp Leu Ile Asp Thr Leu Gly Ser Ala Pro Phe Cys Phe Val
    370                 375                 380

Tyr Glu
385
```

<210> SEQ ID NO 11
<211> LENGTH: 410
<212> TYPE: PRT
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 11

```
Met Leu Val Tyr Ala Ser Leu Ala Ile Leu Pro Leu Leu Ala Gly Val
1               5                   10                  15

Gly Ala Ser Pro Leu Asn Lys Lys Ala Thr Phe Ser Tyr Gln Gln Val
            20                  25                  30

Ser Arg Asn Pro Asp Phe Pro Leu Asp Val Val Asp Glu Leu Glu Ala
        35                  40                  45

Lys Ala Met Pro Gly Val Glu Thr Trp Met Ala Gly Lys Asp Asn Ala
    50                  55                  60

Asn Gly Cys Thr Leu Glu Asn Ala Ala Val Arg Arg Glu Trp Gly Asp
65                  70                  75                  80

Leu Ser Val Ala Glu Arg Glu Glu Tyr Val Ala Ala Val Leu Cys Leu
                85                  90                  95

Gln Lys Leu Pro Ser Lys Ala Pro Glu Gly Lys Ala Pro Gly Ala Leu
            100                 105                 110
```

-continued

```
Ser Arg Phe Asp Asp Phe Val Ala Thr His Met Thr Gln Ala Met Met
        115                 120                 125

Leu His Ser Pro Thr Asn Leu Phe Ala Ser His Arg Tyr Tyr Ile Trp
    130                 135                 140

Ala Tyr Glu Thr Ala Leu Arg Glu Glu Cys Gly Tyr Thr Gly Tyr Gln
145                 150                 155                 160

Pro Tyr Met Asn Tyr Asp Arg Tyr Ala Asp Asp Leu Leu Asn Ser Pro
                165                 170                 175

Leu Phe Asn Gly Asn Ala Ser Ser Leu Gly Gly Asn Gly Ala Pro Ser
            180                 185                 190

Gln Tyr Ala Gly Val Pro Gln Pro Phe Arg Pro Pro Tyr Asn Met Ile
        195                 200                 205

Pro Ser Ala Gly Gly Gly Gly Cys Val Thr Glu Gly Pro Phe Lys Asp
    210                 215                 220

Met Val Val Ser Leu Gly Pro Val Gly Ile Ile Val Asn Asp Ile Pro
225                 230                 235                 240

Lys Asn Pro Arg Ser Asp Gly Leu Gly Ser Asn Pro Arg Cys Leu Arg
                245                 250                 255

Arg Asp Val Asn Lys Phe Ser Ala Ala Gly Ala Lys Ala Asn Tyr Thr
            260                 265                 270

Tyr Asn Leu Ile Thr Glu Asn Ala Gly Ile Asp Ala Phe Tyr Asn Arg
        275                 280                 285

Tyr Leu Gly Gln Pro Gln Leu Lys Asp Asp Pro Asn Pro Trp Gly Leu
    290                 295                 300

His Asn Ala Gly His Tyr Ile Val Gly Gly Asp Pro Gly Gly Asp Phe
305                 310                 315                 320

Tyr Ala Ser Pro Gly Asp Pro Tyr Phe Tyr Phe His His Gly Met Leu
                325                 330                 335

Asp Arg Val Trp Trp Ile Trp Gln Met Gln Asp Pro Ala Ala Arg Val
            340                 345                 350

Asn Leu Ile Pro Gly Thr Gly Ala Pro Ala Met Asn His Pro Gly Met
        355                 360                 365

Pro Met Asn Arg Arg Gln Ser Ser Ala Thr Ile Val Asp Leu Gly Trp
    370                 375                 380

Thr Ala Pro Ala Val Pro Ile Thr Glu Leu Asn Asp Ser Leu Gly Gly
385                 390                 395                 400

Asn Gly Gly Lys Phe Cys Tyr Val Tyr Val
                405                 410


<210> SEQ ID NO 12
<211> LENGTH: 384
<212> TYPE: PRT
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 12

Met Arg Leu Phe Lys Ala Leu Ala Ala Ala Ala Leu Thr Gly Leu Val
1               5                   10                  15
```

-continued

```
Ala Ala Asp Ala Val Asn Asp Leu Glu Thr Lys Gly Arg Ala Ala Leu
            20                  25                  30

Asp Ala Val Ile Glu Ser Ser Thr Thr Cys Ser Lys Asp Lys Leu Lys
        35                  40                  45

Val Arg Arg Glu Trp Gly Asp Ile Ser Thr Thr Glu Arg Lys Ala Tyr
    50                  55                  60

Leu Asp Gly Val Leu Cys Leu Leu Asn Thr Pro Ser Lys Leu Asp Pro
65                  70                  75                  80

Ala Arg Tyr Pro Gly Ala Lys Asn Arg Tyr Asp Asp Phe Val Val Val
                85                  90                  95

His Met Asn Gln Thr Leu Ser Ile His Gly Thr Gly Asn Phe Leu Val
            100                 105                 110

Trp His Arg Tyr Tyr Val Trp Ala Trp Glu Asn Val Met Arg Thr Glu
        115                 120                 125

Cys Gly Tyr Glu Gly Thr Gln Pro Tyr Trp Asp Tyr Gly Arg Trp Ala
    130                 135                 140

Glu Asp Pro Leu Ser Ser Pro Leu Phe Asp Gly Ser Glu Thr Ser Leu
145                 150                 155                 160

Gly Gly Asn Gly Ala Pro Val Thr Gln Asn Lys Arg Ser Arg Val Glu
                165                 170                 175

Gly Arg Gln Phe Gly Gly Gly Gly Gly Phe Gly Gly Gly Leu Gly Gly
            180                 185                 190

Gly Phe Gly Gly Gly Gly Asp Gly Gly Gly Cys Ile Ser Thr Gly Pro
        195                 200                 205

Phe Lys Asp Met Val Val Thr Leu Gly Pro Met Ser Ala Val Val Arg
    210                 215                 220

Pro Ala Pro Ala Arg Asn Pro Gln Ala Asp Gly Tyr Gly Ser Asn Pro
225                 230                 235                 240

Arg Cys Ile Arg Arg Asp Ile Thr Asn Ser Leu Ser Met Ala Tyr Gly
                245                 250                 255

Lys Thr Glu Asp Ile Val Asn Ser Ile Val Asn Tyr Asn Asp Ile Leu
            260                 265                 270

Ala Phe Gln Asn Phe Met Gln Gly Gly Thr Gly Val His Gly Val Gly
        275                 280                 285

His Phe Thr Val Ser Gly Asp Pro Gly Gly Asp Phe Tyr Ile Ser Pro
    290                 295                 300

Asn Glu Pro Ser Phe Trp Leu His His Ala Met Ile Asp Arg Ile Trp
305                 310                 315                 320

Thr Ile Trp Gln Ser Gln Asp Tyr Glu Thr Arg Arg Gly Ala Met Glu
                325                 330                 335

Gly Gly Thr Ser Met Met Gly Gly Gly Arg Ala Gln Ser Leu Asp Asp
            340                 345                 350

Leu Val Asp Leu Gly Val Ile Ala Asp Thr Val Tyr Pro Ile Arg Asp
        355                 360                 365

Ile Leu Ser Ser Val Asp Gly Pro Gly Pro Phe Cys Tyr Val Tyr Glu
    370                 375                 380
```

-continued

<210> SEQ ID NO 13
<211> LENGTH: 395
<212> TYPE: PRT
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 13

Met Arg Ser Ala Leu Ala Leu Val Leu Ala Ala Ser Leu Leu Gly Gly
1 5 10 15

Glu Ala Ser Ser Ile Lys Lys Arg Phe Ser Thr Leu Asp Val Trp Arg
20 25 30

His Gly Asp Tyr Glu Arg Asp Ile Val Asp Gln Leu Ser Asp Glu Thr
35 40 45

Phe Pro Lys Ile Ala Glu Trp Val Glu Lys Thr Gly Ser Thr Cys Thr
50 55 60

Leu Glu Asn Ala Val Gln Arg Lys Glu Trp Thr Asp Leu Thr Ile Asp
65 70 75 80

Glu Arg Ala Asp Tyr Ile Gln Ala Val Gln Cys Leu Met Lys Leu Pro
85 90 95

Pro Lys Ser Gln Asp Gln Val Pro Gly Ser Leu Asn Arg Tyr Asp Asp
100 105 110

Phe Val Ala Thr His Val Thr Gly Ile Pro Val Leu His Ala Pro Thr
115 120 125

Asn Leu Phe Ala Ser His Arg Tyr Tyr Ile Trp Ala Tyr Glu Leu Ala
130 135 140

Leu Arg Glu Glu Cys Gly Tyr Lys Gly Tyr Gln Pro Tyr Met Asn Tyr
145 150 155 160

Glu Arg His Gln Asp Pro Ile Thr Ser Pro Leu Phe Asn Gly Asn Ala
165 170 175

Thr Ser Met Gly Gly Asn Gly Ala Ala Ala Glu Tyr Pro Gly Val Val
180 185 190

Met Pro Tyr Pro Arg Pro Tyr Asn Val Ile Pro Ala Ala Gly Gly Gly
195 200 205

Gly Cys Val Thr Glu Gly Pro Phe Ser Asp Met Val Val Ser Ile Gly
210 215 220

Pro Leu Gly Thr Val Leu Arg Asp Ile Pro Arg Asn Pro Arg Ala Asp
225 230 235 240

Gly Leu Gly Ser Asn Pro Arg Cys Leu Arg Arg Asp Leu Asn Lys Phe
245 250 255

Ser Ala Ala Gly Ala Ser Ala Asn His Ser Tyr Ser Leu Ile Met Asp
260 265 270

Tyr Pro Asp Val Asp Ala Phe Tyr Asn Arg Tyr Leu Gly Gln Pro Phe
275 280 285

Leu Arg Gly Asp Glu Phe Pro Trp Gly Leu His Ser Ala Gly His Tyr
290 295 300

Ile Thr Gly Gly Asp Pro Gly Gly Asp Phe Tyr Ala Ser Pro Gly Asp
305 310 315 320

Pro Thr Phe Trp Met His His Ala Ala Leu Asp Arg Leu Trp Trp Leu
325 330 335

-continued

```
Trp Gln Met Gln Asp Pro Glu Asn Arg Leu Gln Ala Ile Pro Gly Ile
            340                 345                 350

Thr Ser Ser Arg Met Thr Asn Glu Asp Ala Gln Lys Thr Met Val Asp
        355                 360                 365

Leu Lys Trp Thr Ala Glu Pro Arg Ser Leu Gly Asp Leu Asn Asp Gln
    370                 375                 380

Met Gly Ser Ala Pro Phe Cys Tyr Ile Tyr Val
385                 390                 395


<210> SEQ ID NO 14
<211> LENGTH: 353
<212> TYPE: PRT
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 14

Met Gln Leu Thr Ile Leu Ala Thr Ala Leu Leu Ala Val Ser Ala Ser
1               5                   10                  15

Ala Ala Pro Thr Cys Cys Thr Asn Ala Glu Arg Arg Glu Trp Arg Thr
            20                  25                  30

Phe Ser Thr Lys Glu Lys Gln Ala Tyr Ile Ala Ala Val Lys Cys Leu
        35                  40                  45

Gln Ser Lys Pro Ser Gln Leu Lys Ser Thr Tyr Pro Thr Ser Gln Asn
    50                  55                  60

Arg Phe Asp Asp Phe Gln Ala Val His Ile Asp Leu Thr Glu Lys Tyr
65                  70                  75                  80

His Phe Thr Gly Pro Phe Gln Ala Trp His Arg Val Phe Leu His Lys
                85                  90                  95

Tyr Glu Ser Asp Leu Arg Gly Leu Cys Ala Tyr Lys Gly Tyr Gln Pro
            100                 105                 110

Tyr Trp Asp Trp Thr Lys Asp Ser Gly Ser Glu Ala Ala Phe Leu Ala
        115                 120                 125

Ser Pro Val Phe Asp Ala Val Asn Gly Phe Gly Gly Asn Gly Pro Tyr
    130                 135                 140

Val Asp Thr Ser Asn Phe Pro Val Thr Asn Val Pro Val Lys Ile Pro
145                 150                 155                 160

Asn Lys Thr Gly Gly Gly Cys Val Gln Asp Gly Ala Phe Val Asn Met
                165                 170                 175

Thr Val Thr Leu Gly Pro Gly Pro Ser Leu Glu Ser Asn Pro Arg Cys
            180                 185                 190

Leu Thr Arg Asp Phe Ser Tyr Trp Leu Ile Ser Arg Thr Leu Thr Lys
        195                 200                 205

Ala Val Val Asp Trp Thr Leu Glu Ala Ala Ser Phe Ala Val Phe Asp
    210                 215                 220

Phe Arg Leu Gln Gly Thr Gly Ile Glu Pro Glu Gly Met Thr Val His
225                 230                 235                 240

Ala Gly Gly His Leu Gly Val Gly Gly Asp Ile Gly Glu Ile Gly Asn
                245                 250                 255
```

-continued

```
Met Tyr Ser Ser Pro Gly Asp Pro Leu Phe Tyr Leu His His Ala Asn
            260                 265                 270

Leu Asp Arg Leu Trp Asp Gln Trp Gln Arg Lys Lys Phe Ala Gln Arg
        275                 280                 285

Val Gln Asp Met Thr Gly Pro Asp Thr Met Trp Ala Tyr Pro Phe Asn
    290                 295                 300

Phe Phe Gly Asp Val Pro Tyr Thr Asn Ile Thr Leu Glu Thr Leu Leu
305                 310                 315                 320

Asp Phe Lys Gly Leu Leu Gly Ser Ser Ser Ala Asp Arg Tyr Val Lys
                325                 330                 335

Ile Lys Asp Val Met Asp Ser Gln Gly Pro Asn Leu Cys Val Phe Tyr
            340                 345                 350

Lys

<210> SEQ ID NO 15
<211> LENGTH: 389
<212> TYPE: PRT
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 15

Met Val Ala Ile Ser Tyr Val Leu Thr Ala Leu Ala Val Ala Ile Pro
1               5                   10                  15

Ala Leu Ala Gln Gly Pro Cys Ser Thr Ile Arg Gln Arg Arg Ala Trp
            20                  25                  30

His Thr Leu Ser Asn Ser Glu Lys Arg Ala Tyr Leu Asp Ala Glu Val
        35                  40                  45

Cys Leu Leu Gly Lys Ala Pro Lys Phe Gly Phe Glu Gly Ala Lys Asn
    50                  55                  60

Arg Phe Glu Glu Leu Gln Ala Ala His Gln Val Gln Ala Tyr Leu Ile
65                  70                  75                  80

His Gly Val Gly Ala Phe Leu Pro Phe His Arg Tyr Leu Met His Ala
                85                  90                  95

His Glu Thr Leu Leu Arg Thr Glu Cys Gly Tyr Gln Gly Ala Gln Pro
            100                 105                 110

Tyr Trp Asp Glu Thr Arg Asp Ala Gly Arg Val Ser Glu Ser Glu Ile
        115                 120                 125

Leu Asp Pro Asp Thr Gly Phe Gly Gly Asp Gly Val Gly Glu Arg Gly
    130                 135                 140

Cys Ile Ala Asp Gly Pro Phe Ala Gly Tyr Ile Asn Ser Ile Gly Pro
145                 150                 155                 160

Gly Tyr Arg Ile Thr Asp Arg Cys Ile Thr Arg Phe Val Asn Asn Thr
                165                 170                 175

Arg Ser Leu Met Ala Ser Pro Arg Phe Thr Asp Arg Cys Gln Gly Met
            180                 185                 190

Asn Arg Tyr Val Asp Val Trp Pro Cys Leu Glu Gly Asn Pro His Asn
        195                 200                 205

Ser Gly His Gly Ala Ile Ser Gly Leu Met Met Asp Pro Ile Ala Ser
    210                 215                 220
```

-continued

```
Pro Gly Asp Pro Ile Phe Tyr Leu His His Thr Trp Leu Asp Lys Leu
225                 230                 235                 240

Trp Trp Glu Trp Gln Ala Met Asp Leu Pro Arg Arg Leu Thr Asp Ile
                245                 250                 255

Gly Gly Arg Asn Thr Gln Asp Gly Ser Glu Gly Phe Pro Gly Ala Pro
            260                 265                 270

Pro Asn Ser Asn Gly Pro Asn Gly Ala Lys Arg Arg Ser Pro Ala Asp
        275                 280                 285

Gly Pro Ile Leu Ile Phe Pro Gly Asp Gly Gly Asn Ile Asp Trp Asp
    290                 295                 300

Glu Ile Asp Trp Glu Lys Ile Gly Phe Pro Gly Leu Gly Gly Gly Asn
305                 310                 315                 320

Gly Gly Pro Ile Gln Leu Pro Pro Gly Val Asp Val Pro Pro Glu Ala
                325                 330                 335

Leu Glu Pro Pro Glu Asp Ala Glu Pro Gln Glu Pro Arg Gly Asp Pro
            340                 345                 350

Gly Asp Val Thr Thr Leu Asn His Val Leu Lys Met Phe Gly Leu Val
        355                 360                 365

Pro Asp Ala Leu Ile Arg Asp Val Met Asp Ile Ala Gly Gly Thr Leu
    370                 375                 380

Cys Tyr Glu Tyr Val
385


<210> SEQ ID NO 16
<211> LENGTH: 348
<212> TYPE: PRT
<213> ORGANISM: Parascedosporium putredinis

<400> SEQUENCE: 16

Met Lys Asn Leu Ala Gly Leu Ile Thr Ala Leu Ala Thr Gly Val Gly
1               5                   10                  15

Met Ala His Thr His Thr Leu Val Pro His Met Gln Asp Ser Thr Pro
            20                  25                  30

Cys Ile Asn Pro Ser Leu Arg Arg Pro Trp Gln Ile Leu Ser Asp Gly
        35                  40                  45

Glu Lys Arg Ser Tyr Leu Asp Ala Gln Leu Cys Val Met Arg Thr Pro
    50                  55                  60

Gln Thr Leu Gly Leu Pro Gly Ala Arg Thr Arg Phe Glu Glu Leu Ala
65                  70                  75                  80

Ala Thr His Gln Ile Gly Ala Arg Ala Ser His Ala Thr Gly Thr Phe
                85                  90                  95

Phe Pro Tyr His Arg Tyr Leu Leu His Ala His Glu Ser Leu Leu Lys
            100                 105                 110

Glu Cys Gly Tyr His Ala Gly Leu Pro Tyr Trp Asp Glu Thr Arg Glu
        115                 120                 125

Ala Gly Asn Phe Ile Lys Ser Thr Ile Phe Glu Ser Gly Leu Gly Phe
    130                 135                 140
```

-continued

```
Gly Gly Phe Gly Ser Asp Leu Lys Gly Cys Ile Glu Asp Gly Pro Phe
145                 150                 155                 160

Ala Asn Leu Thr Ser Thr Ile Gly Pro Gly Phe Ser Leu Asn Glu His
                165                 170                 175

Cys Ile Ser Arg Ala Leu Asn Glu Thr Ala Gly Leu Lys Ala Ala Arg
            180                 185                 190

Glu Glu Val Asp Lys Cys Leu Glu Ala Asn Asp Tyr Thr Glu Met Trp
        195                 200                 205

Arg Cys Ala Tyr Thr Thr Pro His Arg Gly Gly His Gly Gly Val Gly
    210                 215                 220

Gly Thr Met Gly Asp Ala Leu Ala Ser Pro Gly Asp Pro Val Phe Tyr
225                 230                 235                 240

Val His His Ala Trp Val Asp Lys Ile Trp Trp Asp Trp Gln Glu Ala
                245                 250                 255

Asp Leu Asp Asn Arg Met Tyr Ala Ile Gly Gly Pro Ser Phe Gln Ser
            260                 265                 270

Pro Asp Ile Gly Phe Pro Glu Val Pro Gly Asp Val Glu Glu Glu Glu
        275                 280                 285

Ala Asn Ile Phe Gly Lys Pro Ser Glu Ala Ile Arg Arg Leu Gln Glu
    290                 295                 300

Leu Trp Ser Ser Ser Asp Pro Ser Arg Glu Thr Thr Leu Glu His Asn
305                 310                 315                 320

Leu Thr Leu Leu Gly Ile Ile Pro Asp Ile Asn Ile Ser Lys Val Met
                325                 330                 335

Asp Thr Arg Gly Gly Tyr Leu Cys Tyr Glu Tyr Val
            340                 345
```

<210> SEQ ID NO 17
<211> LENGTH: 1158
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: codon optimised sequence

<400> SEQUENCE: 17

```
caagaaccgg cactgaccga agatgatttt agcattccgg aaattgaagg tggtgatgca       60

ctggcacagc tggcccaact ggcagcagat agcagccaag aaaccgcact gcgtatggca      120

aaacgtggtc tgaatagcgg ttgtagcccg agccagatta aagttcgtcg tgaatggcgt      180

accctgacca gcgcacagcg taaacagtat atcgcaagcg ttaaatgcct gcagaccaaa      240

ccgagctttt ttgatccgaa cattattccg gcagcaaaaa gcctgtttga tgattttgtt      300

ggtgtgcatg tttttcagac cggcagcatt catctgaccg caacctttct gacctggcat      360

cgttattttg tgtataccta tgaaaccaaa ctgcgcgaag aatgtggtta tacaggtccg      420

ctgccgtatt gggaatgggg tttagatgtt aataatccga atgccagtcc ggtttttgat      480

ggtagcgata ccagcctgag cggtaatggt gcatttttttg cacatgaagg tattcagatg      540

gtgcagccga ttaatggcaa tattctgaaa ctgcctcctg gtaatggcgg tggttgtgtt      600
```

-continued

accaaaggtc cgtttaaaga tatgaaagtg catttcggca ccattattct gccggtttat 660

ggccagccga ttctgagcgg tgttgaaaat ccgattgcag ataatgaacg ttgcctgaaa 720

cgcgatctga atgcaggtat tgccaaacgt tttaccagct ttctgaatag taccagcgtg 780

attctgaaga acaacaacat cgaaatgttt caggcccatc tgcagggtga tgatcgttat 840

gttctgaatc agctgggtgt tcatggtggt ggtcattata ccattggtgg tgatcctggt 900

ggcgatccgt ttattagtcc gggtgatccg gcattttatc tgcatcatgc acagattgat 960

cgcatctatt ggatttggca gatgctggat ttcaaaaatc gtcagggcgt tcatggtaca 1020

gcaacactgc agaataaccc tccgagcgca aatgttaccg ttgaagatac cattgatctg 1080

agtccgctgg caccgcctgt taaaatcaaa gatctgatga ataccgttgg tggcagtccg 1140

ctgtgttata tctatctg 1158

<210> SEQ ID NO 18
<211> LENGTH: 1161
<212> TYPE: DNA
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 18

caggaacctg ccctcaccga ggatgatttc agcatccccg agatcgaggg aggtgatgcc 60

ctcgcacagc ttgcccagct cgcggcggat tcttctcagg agactgccct caggatggcg 120

aagcgcggtc tgaacagcgg ctgcagccct agccagatca aggttagaag ggaatggcga 180

accctgactt ctgctcagcg caagcagtac attgcgtcgg tcaagtgtct tcagacgaag 240

cccagcttct tcgaccccaa catcatcccg gccgctaagt cgctatttga cgactttgtt 300

ggtgtccatg ttttccagac tgggtccatc catctcactg ccactttcct cacatggcat 360

cgctacttcg tctataccta cgagacgaag ctccgcgagg aatgcggtta cactggaccc 420

ttgccctact gggagtgggg actagacgtc aacaacccca acgcctcccc ggtcttcgac 480

ggctccgaca cgtctctgag cggcaacggt gccttctttg cgcatgaggg catccagatg 540

gtgcagccta tcaacggcaa catcctcaag ctccccccg gcaacggtgg tggctgcgtg 600

accaagggtc ccttcaagga catgaaggtt cactttggca ccatcatcct gcccgtgtac 660

ggccagccta tcctcagtgg tgtcgagaac cccattgccg acaacgagcg ctgcctcaag 720

cgcgatctca atgccggcat cgctaagcgc ttcactagct tcctcaactc gaccagcgtg 780

attctcaaga acaacaacat cgagatgttc caggcccatc tgcagggcga cgaccgctac 840

gtgctcaacc agctcggtgt tcacggtggt ggtcactaca ccatcggcgg tgaccccggt 900

ggtgatccct tcatctcccc tggtgacccg gctttctacc tccaccacgc ccagattgac 960

cgcatctact ggatctggca gatgctcgac ttcaagaacc gtcagggtgt ccacggtacc 1020

gccaccctcc agaacaaccc tcccagcgcc aacgttaccg ttgaggacac catcgacctc 1080

tctcccctcg ccccgcctgt caagatcaag gacctcatga acactgtcgg aggctcgccc 1140

ttgtgctaca tctacctcta a 1161

<210> SEQ ID NO 19

-continued

<211> LENGTH: 1092
<212> TYPE: DNA
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 19

cagacctatg agtttagcga ggaggagctc acctctggtg atgccctcaa ggcccttagc 60

aagcaggcta tggaaaatgc tcttgcccgc ctccccgaat ctggagaggg ctgcacgcgg 120

gagaacgtca aaattcgaaa ggaatggcgc aacatgccgg ccgagatgag aatcggctat 180

gtcagcgctc tccagtgctt gatggaatcc gaaagcgaat atcccgacgt cgacggtgcc 240

aagacggcgt ttgacgactt cgccgttctt cattacaacc tcacgccgtt cgtgcataac 300

tctgctacct tccttacctt ccacaggtac tacattcaca ccctggaaga gcagatgagg 360

aacaagtgcg gatacactgg tgacttcccc tactgggagt ggggcctcga ctgcgacgac 420

ccgcaacagt ctcccctctt tgacggctcc gaaacctctc tcggcagcga cggtgagcct 480

gtggaggccg gagccggcgg tggcttcggc ggcggcttcg gctttggcat gggaggcggc 540

agcggtggtg gctgcgtgat gaagggaccc ttctccaact acaccgtcaa cctcggaccc 600

tcaaccaccg ccgacccgct cgcatataac ccgcgctgca tcaagcgaaa cctgaacggc 660

gctatctgca agcagaatgc ctcgctccgg aacacgacga cgacgatcct tgactcgccc 720

gatatcgaac tattccaggc gattgtccaa ggtgacatgc gataccccga ggcgaggggt 780

cttggcatgg ccgtccacgg cggtggtcac tttactattg gtggtgatcc cggtggcgac 840

ttctacttct cccctctcga gcccgccttc ttccaacacc acggccagat cgaccgcatg 900

tactttgtct ggcagaacct cgactgggaa acccggcaga acattgccgg caccggtacc 960

atgatgaacc agccccccag cccagaggtt gagatcaccg aactcctcga cctcagcccc 1020

cttgcagaag ccaggccgat caaggacctc attgacaccc ttggctcggc cccgttctgc 1080

tttgtttacg ag 1092

<210> SEQ ID NO 20
<211> LENGTH: 1176
<212> TYPE: DNA
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 20

tcgccgctta acaagaaggc cacctttcg taccagcagg tatctcgaaa tcctgacttc 60

ccgttggatg tggtcgacga gctcgaggcc aaagccatgc cgggggttga gacatggatg 120

gctgggaagg acaacgccaa cggctgcacg cttgaaaatg ctgcggtccg tagagaatgg 180

ggtgacttat ccgttgccga gcgtgaagag tacgttgccg cggttctctg tctccagaag 240

ttaccatcca aggcgcccga gggaaaggca ccaggtgcac tcagccgctt cgatgacttc 300

gttgccacgc acatgacaca ggccatgatg cttcactcgc ccaccaatct gtttgcgagc 360

caccgttact atatctgggc ctacgaaacc gctcttcgtg aagagtgcgg ctatacgggc 420

taccagccgt acatgaatta cgaccgttat gccgatgacc tcctcaactc ccccttgttc 480

aacgggaacg cgtccagcct aggaggcaac ggagctccaa gccaatacgc cggtgtccct 540

-continued

cagccattca ggcctccgta caacatgatc ccctctgctg gaggcggtgg ttgtgtgacc 600

gagggtccat tcaaagacat ggtcgtgagt cttggccccg tgggaatcat cgtcaatgat 660

atccccaaga acccgcgctc cgatggcctc ggctccaacc cgcgatgcct ccgtcgggac 720

gtgaataaat tttcggcggc cggcgccaag gccaactaca cgtacaacct gattacggag 780

aacgccggaa ttgatgcgtt ttataaccgc tacctcgggc agccgcagct gaaggatgat 840

cccaatcctt gggggctcca taatgccggc cactatatcg taggcggcga cccaggcggt 900

gatttctacg cttcccccgg cgacccctac ttctacttcc accacggcat gttggaccgc 960

gtgtggtgga tctggcagat gcaggacccg gccgcccgcg tcaaccttat tccgggtacc 1020

ggcgccccgg cgatgaacca tcccggcatg ccgatgaacc gtcggcagtc gtccgcgaca 1080

attgtggact tgggttggac ggcgccggcg gtgcccatta cggagctcaa cgattccttg 1140

ggcggtaacg gcggcaagtt ttgctacgtg tacgtg 1176

<210> SEQ ID NO 21
<211> LENGTH: 1098
<212> TYPE: DNA
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 21

gatgctgtca atgacctaga gaccaagggc cgtgccgccc tcgacgccgt cattgagagc 60

tctaccacgt gtagcaagga caagctcaag gtccgcagag agtggggaga tatcagcacc 120

accgagcgca aggcttacct cgacggagtg ctgtgcctcc tgaacacacc ctccaagctc 180

gatcccgctc gttaccccgg cgccaagaac cgctacgatg actttgtcgt tgttcacatg 240

aaccagaccc tctccatcca tggaaccggt aacttcctcg tgtggcaccg ctactacgtg 300

tgggcctggg agaacgtcat gaggaccgag tgtggctatg aaggaaccca gccctactgg 360

gactacggcc gctgggccga ggatcctctt tcctcgcctc tcttcgacgg aagcgagact 420

tcgcttggcg gaaacggcgc ccccgtaacc cagaacaagc gcagccgcgt ggaaggccgc 480

cagttcggcg gtggcggtgg tttcggtggt ggccttggcg gaggtttcgg aggcggcggt 540

gatggcggtg gctgcatttc caccggtccc ttcaaggaca tggtcgtcac cctcggcccc 600

atgtccgccg tcgtcaggcc cgcaccggcc cgcaacccccc aggccgacgg ctacggtagt 660

aacccccgct gcatccgccg cgacatcacc aactcgctga gcatggccta cggaaagacc 720

gaggacatcg tcaacagcat cgtcaactac aacgacatcc ttgccttcca gaacttcatg 780

cagggcggta ccggcgtgca cggcgtcggc cacttcaccg tctccggcga ccccggtggt 840

gatttctaca tctcccccaa cgagccttcc ttctggctcc accacgccat gattgaccgc 900

atctggacca tctggcagtc ccaggactac gagacccgca ggggcgccat ggagggaggc 960

accagcatga tgggaggtgg cagggcccag tccctcgatg acctcgtcga cctcggcgtc 1020

attgccgaca ctgtctaccc catccgcgac atcctcagct ctgttgacgg ccccggcccc 1080

ttctgctacg tgtacgag 1098

<210> SEQ ID NO 22

-continued

<211> LENGTH: 1131
<212> TYPE: DNA
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 22

agcagcatca agaagcgatt ctcgacactc gatgtttggc gccacggcga ctacgagcgg 60

gatatcgtcg atcagctctc cgacgaaacg tttcccaaga tcgccgagtg ggtcgagaag 120

accggctcga cctgcactct cgaaaatgcc gtgcagcgaa aggagtggac cgatttgacc 180

attgatgaga gggcggacta catccaagcg gtgcagtgct tgatgaagct ccctcccaag 240

tcgcaggacc aagttcccgg ctccctcaac cgatatgacg acttcgtggc cactcacgtt 300

actggtattc cggttcttca cgcacctacc aacctcttcg ccagccacag gtactatatc 360

tgggcgtacg agttggcact tcgcgaggag tgcgggtaca agggctacca gccgtacatg 420

aactacgagc gacaccagga tcccatcacc tcgcccctgt tcaacggaaa tgccaccagc 480

atgggcggca atggagcggc ggctgagtac cccggcgtag tcatgcctta tcccaggccc 540

tacaacgtca ttcccgctgc aggcggtggt ggctgcgtca cggaaggtcc cttctccgac 600

atggttgtca gcatcggccc tctgggcact gttctccgcg acattccccg caacccccgc 660

gccgatggcc tcggttccaa cccccgctgc ctgcgacgcg atcttaacaa gttctccgcc 720

gctggagcct ccgccaacca ctcgtactcg ctcattatgg actaccccga tgtcgacgcc 780

ttctacaacc gatacctcgg acagccgttc ctgagaggag atgaattccc gtggggtctt 840

cactctgccg gtcactacat cacgggagga gacccaggtg gcgactttta cgcctcgcct 900

ggtgacccga ccttctggat gcatcatgcc gccctcgacc gcttgtggtg gctgtggcag 960

atgcaggatc ccgagaaccg cctgcaggcc atccccggca tcacttcgtc gaggatgacc 1020

aacgaggatg ctcaaaagac aatggtggat ctgaagtgga cggcagagcc ccgctcgctc 1080

ggagatctca acgatcaaat gggaagtgcc cccttctgtt atatctatgt a 1131

<210> SEQ ID NO 23
<211> LENGTH: 1008
<212> TYPE: DNA
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 23

gcgcccacct gctgcaccaa cgccgagcgg cgagagtgga gaactttctc aactaaggag 60

aagcaagcct acatcgcggc agtgaaatgt cttcaaagca agccatctca gttgaagagc 120

acctacccga cgtcgcagaa tcgattcgat gactttcagg cggttcatat tgacctgacg 180

gagaagtatc actttactgg acctttccag gcctggcatc gcgtctttct ccacaagtac 240

gaatccgacc tccggggact ttgtgcatac aagggctacc agccatactg ggactggacc 300

aaggattccg ggtctgaggc tgccttcctc gcctcgcctg tctttgacgc tgttaatggc 360

ttcggtggaa acggacctta cgttgatacg tcgaactttc ccgtcaccaa cgtccccgtc 420

aaaatcccga ataagactgg cggaggctgt gttcaagatg gcgcctttgt caatatgaca 480

gtcactctcg gcccaggccc tagtcttgag tcgaaccccc gctgtctcac ccgtgacttc 540

-continued

agctactggc tcatctcacg gaccctaacc aaggcggtgg tagactggac tcttgaggcc 600

gcgtctttcg ctgtctttga cttccgtctg cagggcacag gcatcgagcc agaaggcatg 660

acagttcatg caggtggcca ccttggtgtt ggtggtgata tcggcgagat cggcaacatg 720

tattcatccc ccggagaccc gctcttctat cttcaccacg ccaatcttga cagactctgg 780

gaccagtggc agaggaagaa gtttgctcag cgtgtccagg atatgaccgg accggataca 840

atgtgggcct atccattcaa cttctttggc gacgtaccct ataccaatat cactctggaa 900

accttgctcg acttcaaggg tctccttggg tccagctcgg cagaccgata tgtcaagatc 960

aaggacgtta tggacagcca gggtccgaac ctctgcgtct tctacaaa 1008

<210> SEQ ID NO 24
<211> LENGTH: 1110
<212> TYPE: DNA
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 24

cagggtccgt gttccacaat tcgtcaacga cgggcatggc atacgctgag caacagcgag 60

aaacgcgctt acttggatgc cgaggtttgc ttgctgggca aggcccccaa gtttggcttc 120

gaggggcca aaaacaggtt cgaggagctc caggccgctc atcaagtcca ggcttatctc 180

attcatggag ttggagcgtt tttgcctttc cacagatatc ttatgcacgc tcatgagacc 240

ctgttgagga ctgagtgtgg gtaccaggga gcgcaaccat actgggatga gacacgggat 300

gccggtcgcg tcagtgagtc cgagatcctt gatccggaca ctggtttcgg aggagatggt 360

gtgggtgaga ggggttgtat cgccgacgga ccgtttgcag gctacatcaa cagcatcggg 420

ccaggttata ggattacgga tcgctgcatc acacgcttcg tgaacaacac ccgaagcttg 480

atggctagcc cccgtttcac cgacaggtgc caaggcatga atcgatatgt tgacgtgtgg 540

ccatgcctcg agggaaaccc tcataattcg ggccatggcg ctattagcgg actgatgatg 600

gacccgatcg ccagccccgg cgaccccatc ttctaccttc atcacacatg gctcgataag 660

ctgtggtggg agtggcaggc catggacctc ccccgacgcc tcaccgacat tggcggccgt 720

aacacccagg acggctccga aggtttcccc ggtgcaccgc ccaactctaa cggtcccaac 780

ggcgcgaagc gacgatcccc cgccgacggc cccatcctca tcttccccgg ggacggagga 840

aacattgact gggacgagat tgactgggaa aagattggct tccccggcct cggcggtgga 900

aatggaggcc cgatccagtt gccgcccggc gttgacgttc ccccggaagc cctggaacct 960

cccgaggatg ccgagcctca agagcccagg ggtgacccgg gcgatgtgac cacgttgaac 1020

cacgtcctga agatgtttgg cctcgttcca gacgccttga tccgagacgt catggatatc 1080

gccggaggca ctctgtgtta cgagtacgtc 1110

<210> SEQ ID NO 25
<211> LENGTH: 990
<212> TYPE: DNA
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 25

-continued

```
catacgcata cacttgtgcc gcacatgcaa gactcaactc cctgtatcaa cccaagtttg       60

cgacgtccat ggcagattct ctcggacggc gaaaaacgct catatctcga tgcccagctt      120

tgtgtgatga gaacgccgca gaccctcggt cttcctggcg cgagaacacg tttcgaggag      180

ctggctgcca cccaccagat tggcgcccgt gccagccatg ccacggggac atttttcccc      240

taccatcgat acctactgca tgcccatgag tcattgctga aggagtgcgg ctaccacgca      300

ggtctccctt actgggatga gaccagggaa gctggaaatt tcatcaagtc taccatattc      360

gaatcgggcc tcggattcgg tggctttgga agcgacctca aagggtgcat cgaagacgga      420

cctttcgcaa acttgacaag tacaatcggt cccgggtttt cgctgaacga acactgcatc      480

tcacgtgcgc tcaatgaaac tgcagggctc aaggcggcta gggaagaggt tgataagtgc      540

ttagaagcca acgactatac agagatgtgg cgttgtgcat ataccacacc ccatcgtggg      600

ggtcatgggg gcgtgggagg cacgatggga gacgctttgg catcgcccgg cgacccggta      660

ttctacgtcc accacgcttg ggtcgataag atttggtggg attggcagga ggctgatctt      720

gataatagaa tgtatgctat tggcgggccc agcttccagt cacctgatat cgggtttcct      780

gaggttcctg gtgatgttga ggaagaagaa gcaaatatct ttggcaaacc aagcgaagca      840

atccgacgac tacaggagct atggagttcc tctgatccaa gcagagaaac taccctggag      900

cataatctga cattgctagg tattattccc gacatcaaca ttagtaaggt catggacact      960

agaggtgggt atctttgcta cgaatatgtt                                       990
```

<210> SEQ ID NO 26
<211> LENGTH: 386
<212> TYPE: PRT
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 26

```
Gln Glu Pro Ala Leu Thr Glu Asp Asp Phe Ser Ile Pro Glu Ile Glu
1               5                  10                  15

Gly Gly Asp Ala Leu Ala Gln Leu Ala Gln Leu Ala Ala Asp Ser Ser
            20                  25                  30

Gln Glu Thr Ala Leu Arg Met Ala Lys Arg Gly Leu Asn Ser Gly Cys
        35                  40                  45

Ser Pro Ser Gln Ile Lys Val Arg Arg Glu Trp Arg Thr Leu Thr Ser
    50                  55                  60

Ala Gln Arg Lys Gln Tyr Ile Ala Ser Val Lys Cys Leu Gln Thr Lys
65                  70                  75                  80

Pro Ser Phe Phe Asp Pro Asn Ile Ile Pro Ala Ala Lys Ser Leu Phe
                85                  90                  95

Asp Asp Phe Val Gly Val His Val Phe Gln Thr Gly Ser Ile His Leu
            100                 105                 110

Thr Ala Thr Phe Leu Thr Trp His Arg Tyr Phe Val Tyr Thr Tyr Glu
        115                 120                 125

Thr Lys Leu Arg Glu Glu Cys Gly Tyr Thr Gly Pro Leu Pro Tyr Trp
    130                 135                 140
```

-continued

```
Glu Trp Gly Leu Asp Val Asn Asn Pro Asn Ala Ser Pro Val Phe Asp
145                 150                 155                 160

Gly Ser Asp Thr Ser Leu Ser Gly Asn Gly Ala Phe Phe Ala His Glu
                165                 170                 175

Gly Ile Gln Met Val Gln Pro Ile Asn Gly Asn Ile Leu Lys Leu Pro
            180                 185                 190

Pro Gly Asn Gly Gly Gly Cys Val Thr Lys Gly Pro Phe Lys Asp Met
        195                 200                 205

Lys Val His Phe Gly Thr Ile Ile Leu Pro Val Tyr Gly Gln Pro Ile
    210                 215                 220

Leu Ser Gly Val Glu Asn Pro Ile Ala Asp Asn Glu Arg Cys Leu Lys
225                 230                 235                 240

Arg Asp Leu Asn Ala Gly Ile Ala Lys Arg Phe Thr Ser Phe Leu Asn
                245                 250                 255

Ser Thr Ser Val Ile Leu Lys Asn Asn Asn Ile Glu Met Phe Gln Ala
            260                 265                 270

His Leu Gln Gly Asp Asp Arg Tyr Val Leu Asn Gln Leu Gly Val His
        275                 280                 285

Gly Gly Gly His Tyr Thr Ile Gly Gly Asp Pro Gly Gly Asp Pro Phe
    290                 295                 300

Ile Ser Pro Gly Asp Pro Ala Phe Tyr Leu His His Ala Gln Ile Asp
305                 310                 315                 320

Arg Ile Tyr Trp Ile Trp Gln Met Leu Asp Phe Lys Asn Arg Gln Gly
                325                 330                 335

Val His Gly Thr Ala Thr Leu Gln Asn Asn Pro Pro Ser Ala Asn Val
            340                 345                 350

Thr Val Glu Asp Thr Ile Asp Leu Ser Pro Leu Ala Pro Pro Val Lys
        355                 360                 365

Ile Lys Asp Leu Met Asn Thr Val Gly Gly Ser Pro Leu Cys Tyr Ile
    370                 375                 380

Tyr Leu
385


<210> SEQ ID NO 27
<211> LENGTH: 364
<212> TYPE: PRT
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 27

Gln Thr Tyr Glu Phe Ser Glu Glu Glu Leu Thr Ser Gly Asp Ala Leu
1               5                   10                  15

Lys Ala Leu Ser Lys Gln Ala Met Glu Asn Ala Leu Ala Arg Leu Pro
            20                  25                  30

Glu Ser Gly Glu Gly Cys Thr Arg Glu Asn Val Lys Ile Arg Lys Glu
        35                  40                  45

Trp Arg Asn Met Pro Ala Glu Met Arg Ile Gly Tyr Val Ser Ala Leu
    50                  55                  60
```

-continued

```
Gln Cys Leu Met Glu Ser Glu Ser Glu Tyr Pro Asp Val Asp Gly Ala
65                  70                  75                  80

Lys Thr Ala Phe Asp Asp Phe Ala Val Leu His Tyr Asn Leu Thr Pro
                85                  90                  95

Phe Val His Asn Ser Ala Thr Phe Leu Thr Phe His Arg Tyr Tyr Ile
            100                 105                 110

His Thr Leu Glu Glu Gln Met Arg Asn Lys Cys Gly Tyr Thr Gly Asp
        115                 120                 125

Phe Pro Tyr Trp Glu Trp Gly Leu Asp Cys Asp Asp Pro Gln Gln Ser
    130                 135                 140

Pro Leu Phe Asp Gly Ser Glu Thr Ser Leu Gly Ser Asp Gly Glu Pro
145                 150                 155                 160

Val Glu Ala Gly Ala Gly Gly Gly Phe Gly Gly Gly Phe Gly Phe Gly
                165                 170                 175

Met Gly Gly Gly Ser Gly Gly Gly Cys Val Met Lys Gly Pro Phe Ser
            180                 185                 190

Asn Tyr Thr Val Asn Leu Gly Pro Ser Thr Thr Ala Asp Pro Leu Ala
        195                 200                 205

Tyr Asn Pro Arg Cys Ile Lys Arg Asn Leu Asn Gly Ala Ile Cys Lys
    210                 215                 220

Gln Asn Ala Ser Leu Arg Asn Thr Thr Thr Thr Ile Leu Asp Ser Pro
225                 230                 235                 240

Asp Ile Glu Leu Phe Gln Ala Ile Val Gln Gly Asp Met Arg Tyr Pro
                245                 250                 255

Glu Ala Arg Gly Leu Gly Met Ala Val His Gly Gly Gly His Phe Thr
            260                 265                 270

Ile Gly Gly Asp Pro Gly Gly Asp Phe Tyr Phe Ser Pro Leu Glu Pro
        275                 280                 285

Ala Phe Phe Gln His His Gly Gln Ile Asp Arg Met Tyr Phe Val Trp
    290                 295                 300

Gln Asn Leu Asp Trp Glu Thr Arg Gln Asn Ile Ala Gly Thr Gly Thr
305                 310                 315                 320

Met Met Asn Gln Pro Pro Ser Pro Glu Val Glu Ile Thr Glu Leu Leu
                325                 330                 335

Asp Leu Ser Pro Leu Ala Glu Ala Arg Pro Ile Lys Asp Leu Ile Asp
            340                 345                 350

Thr Leu Gly Ser Ala Pro Phe Cys Phe Val Tyr Glu
        355                 360


<210> SEQ ID NO 28
<211> LENGTH: 392
<212> TYPE: PRT
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 28

Ser Pro Leu Asn Lys Lys Ala Thr Phe Ser Tyr Gln Gln Val Ser Arg
1               5                   10                  15
```

-continued

```
Asn Pro Asp Phe Pro Leu Asp Val Val Asp Glu Leu Glu Ala Lys Ala
            20                  25                  30

Met Pro Gly Val Glu Thr Trp Met Ala Gly Lys Asp Asn Ala Asn Gly
        35                  40                  45

Cys Thr Leu Glu Asn Ala Ala Val Arg Arg Glu Trp Gly Asp Leu Ser
    50                  55                  60

Val Ala Glu Arg Glu Glu Tyr Val Ala Ala Val Leu Cys Leu Gln Lys
65                  70                  75                  80

Leu Pro Ser Lys Ala Pro Glu Gly Lys Ala Pro Gly Ala Leu Ser Arg
                85                  90                  95

Phe Asp Asp Phe Val Ala Thr His Met Thr Gln Ala Met Met Leu His
            100                 105                 110

Ser Pro Thr Asn Leu Phe Ala Ser His Arg Tyr Tyr Ile Trp Ala Tyr
        115                 120                 125

Glu Thr Ala Leu Arg Glu Glu Cys Gly Tyr Thr Gly Tyr Gln Pro Tyr
    130                 135                 140

Met Asn Tyr Asp Arg Tyr Ala Asp Asp Leu Leu Asn Ser Pro Leu Phe
145                 150                 155                 160

Asn Gly Asn Ala Ser Ser Leu Gly Gly Asn Gly Ala Pro Ser Gln Tyr
                165                 170                 175

Ala Gly Val Pro Gln Pro Phe Arg Pro Pro Tyr Asn Met Ile Pro Ser
            180                 185                 190

Ala Gly Gly Gly Gly Cys Val Thr Glu Gly Pro Phe Lys Asp Met Val
        195                 200                 205

Val Ser Leu Gly Pro Val Gly Ile Ile Val Asn Asp Ile Pro Lys Asn
    210                 215                 220

Pro Arg Ser Asp Gly Leu Gly Ser Asn Pro Arg Cys Leu Arg Arg Asp
225                 230                 235                 240

Val Asn Lys Phe Ser Ala Ala Gly Ala Lys Ala Asn Tyr Thr Tyr Asn
                245                 250                 255

Leu Ile Thr Glu Asn Ala Gly Ile Asp Ala Phe Tyr Asn Arg Tyr Leu
            260                 265                 270

Gly Gln Pro Gln Leu Lys Asp Asp Pro Asn Pro Trp Gly Leu His Asn
        275                 280                 285

Ala Gly His Tyr Ile Val Gly Gly Asp Pro Gly Gly Asp Phe Tyr Ala
    290                 295                 300

Ser Pro Gly Asp Pro Tyr Phe Tyr Phe His His Gly Met Leu Asp Arg
305                 310                 315                 320

Val Trp Trp Ile Trp Gln Met Gln Asp Pro Ala Ala Arg Val Asn Leu
                325                 330                 335

Ile Pro Gly Thr Gly Ala Pro Ala Met Asn His Pro Gly Met Pro Met
            340                 345                 350

Asn Arg Arg Gln Ser Ser Ala Thr Ile Val Asp Leu Gly Trp Thr Ala
        355                 360                 365

Pro Ala Val Pro Ile Thr Glu Leu Asn Asp Ser Leu Gly Gly Asn Gly
    370                 375                 380
```

-continued

```
Gly Lys Phe Cys Tyr Val Tyr Val
385                 390


<210> SEQ ID NO 29
<211> LENGTH: 366
<212> TYPE: PRT
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 29

Asp Ala Val Asn Asp Leu Glu Thr Lys Gly Arg Ala Ala Leu Asp Ala
1               5                   10                  15

Val Ile Glu Ser Ser Thr Thr Cys Ser Lys Asp Lys Leu Lys Val Arg
            20                  25                  30

Arg Glu Trp Gly Asp Ile Ser Thr Thr Glu Arg Lys Ala Tyr Leu Asp
        35                  40                  45

Gly Val Leu Cys Leu Leu Asn Thr Pro Ser Lys Leu Asp Pro Ala Arg
    50                  55                  60

Tyr Pro Gly Ala Lys Asn Arg Tyr Asp Asp Phe Val Val Val His Met
65                  70                  75                  80

Asn Gln Thr Leu Ser Ile His Gly Thr Gly Asn Phe Leu Val Trp His
                85                  90                  95

Arg Tyr Tyr Val Trp Ala Trp Glu Asn Val Met Arg Thr Glu Cys Gly
            100                 105                 110

Tyr Glu Gly Thr Gln Pro Tyr Trp Asp Tyr Gly Arg Trp Ala Glu Asp
        115                 120                 125

Pro Leu Ser Ser Pro Leu Phe Asp Gly Ser Glu Thr Ser Leu Gly Gly
    130                 135                 140

Asn Gly Ala Pro Val Thr Gln Asn Lys Arg Ser Arg Val Glu Gly Arg
145                 150                 155                 160

Gln Phe Gly Gly Gly Gly Gly Phe Gly Gly Gly Leu Gly Gly Gly Phe
                165                 170                 175

Gly Gly Gly Gly Asp Gly Gly Gly Cys Ile Ser Thr Gly Pro Phe Lys
            180                 185                 190

Asp Met Val Val Thr Leu Gly Pro Met Ser Ala Val Val Arg Pro Ala
        195                 200                 205

Pro Ala Arg Asn Pro Gln Ala Asp Gly Tyr Gly Ser Asn Pro Arg Cys
    210                 215                 220

Ile Arg Arg Asp Ile Thr Asn Ser Leu Ser Met Ala Tyr Gly Lys Thr
225                 230                 235                 240

Glu Asp Ile Val Asn Ser Ile Val Asn Tyr Asn Asp Ile Leu Ala Phe
                245                 250                 255

Gln Asn Phe Met Gln Gly Gly Thr Gly Val His Gly Val Gly His Phe
            260                 265                 270

Thr Val Ser Gly Asp Pro Gly Gly Asp Phe Tyr Ile Ser Pro Asn Glu
        275                 280                 285

Pro Ser Phe Trp Leu His His Ala Met Ile Asp Arg Ile Trp Thr Ile
    290                 295                 300
```

-continued

```
Trp Gln Ser Gln Asp Tyr Glu Thr Arg Arg Gly Ala Met Glu Gly Gly
305                 310                 315                 320

Thr Ser Met Met Gly Gly Gly Arg Ala Gln Ser Leu Asp Asp Leu Val
                325                 330                 335

Asp Leu Gly Val Ile Ala Asp Thr Val Tyr Pro Ile Arg Asp Ile Leu
            340                 345                 350

Ser Ser Val Asp Gly Pro Gly Pro Phe Cys Tyr Val Tyr Glu
        355                 360                 365


<210> SEQ ID NO 30
<211> LENGTH: 377
<212> TYPE: PRT
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 30

Ser Ser Ile Lys Lys Arg Phe Ser Thr Leu Asp Val Trp Arg His Gly
1               5                   10                  15

Asp Tyr Glu Arg Asp Ile Val Asp Gln Leu Ser Asp Glu Thr Phe Pro
            20                  25                  30

Lys Ile Ala Glu Trp Val Glu Lys Thr Gly Ser Thr Cys Thr Leu Glu
        35                  40                  45

Asn Ala Val Gln Arg Lys Glu Trp Thr Asp Leu Thr Ile Asp Glu Arg
    50                  55                  60

Ala Asp Tyr Ile Gln Ala Val Gln Cys Leu Met Lys Leu Pro Pro Lys
65                  70                  75                  80

Ser Gln Asp Gln Val Pro Gly Ser Leu Asn Arg Tyr Asp Asp Phe Val
                85                  90                  95

Ala Thr His Val Thr Gly Ile Pro Val Leu His Ala Pro Thr Asn Leu
            100                 105                 110

Phe Ala Ser His Arg Tyr Tyr Ile Trp Ala Tyr Glu Leu Ala Leu Arg
        115                 120                 125

Glu Glu Cys Gly Tyr Lys Gly Tyr Gln Pro Tyr Met Asn Tyr Glu Arg
    130                 135                 140

His Gln Asp Pro Ile Thr Ser Pro Leu Phe Asn Gly Asn Ala Thr Ser
145                 150                 155                 160

Met Gly Gly Asn Gly Ala Ala Ala Glu Tyr Pro Gly Val Val Met Pro
                165                 170                 175

Tyr Pro Arg Pro Tyr Asn Val Ile Pro Ala Ala Gly Gly Gly Gly Cys
            180                 185                 190

Val Thr Glu Gly Pro Phe Ser Asp Met Val Val Ser Ile Gly Pro Leu
        195                 200                 205

Gly Thr Val Leu Arg Asp Ile Pro Arg Asn Pro Arg Ala Asp Gly Leu
    210                 215                 220

Gly Ser Asn Pro Arg Cys Leu Arg Arg Asp Leu Asn Lys Phe Ser Ala
225                 230                 235                 240

Ala Gly Ala Ser Ala Asn His Ser Tyr Ser Leu Ile Met Asp Tyr Pro
                245                 250                 255
```

-continued

```
Asp Val Asp Ala Phe Tyr Asn Arg Tyr Leu Gly Gln Pro Phe Leu Arg
            260                 265                 270

Gly Asp Glu Phe Pro Trp Gly Leu His Ser Ala Gly His Tyr Ile Thr
        275                 280                 285

Gly Gly Asp Pro Gly Gly Asp Phe Tyr Ala Ser Pro Gly Asp Pro Thr
    290                 295                 300

Phe Trp Met His His Ala Ala Leu Asp Arg Leu Trp Trp Leu Trp Gln
305                 310                 315                 320

Met Gln Asp Pro Glu Asn Arg Leu Gln Ala Ile Pro Gly Ile Thr Ser
                325                 330                 335

Ser Arg Met Thr Asn Glu Asp Ala Gln Lys Thr Met Val Asp Leu Lys
            340                 345                 350

Trp Thr Ala Glu Pro Arg Ser Leu Gly Asp Leu Asn Asp Gln Met Gly
        355                 360                 365

Ser Ala Pro Phe Cys Tyr Ile Tyr Val
    370                 375


<210> SEQ ID NO 31
<211> LENGTH: 336
<212> TYPE: PRT
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 31

Ala Pro Thr Cys Cys Thr Asn Ala Glu Arg Arg Glu Trp Arg Thr Phe
1               5                   10                  15

Ser Thr Lys Glu Lys Gln Ala Tyr Ile Ala Ala Val Lys Cys Leu Gln
            20                  25                  30

Ser Lys Pro Ser Gln Leu Lys Ser Thr Tyr Pro Thr Ser Gln Asn Arg
        35                  40                  45

Phe Asp Asp Phe Gln Ala Val His Ile Asp Leu Thr Glu Lys Tyr His
    50                  55                  60

Phe Thr Gly Pro Phe Gln Ala Trp His Arg Val Phe Leu His Lys Tyr
65                  70                  75                  80

Glu Ser Asp Leu Arg Gly Leu Cys Ala Tyr Lys Gly Tyr Gln Pro Tyr
                85                  90                  95

Trp Asp Trp Thr Lys Asp Ser Gly Ser Glu Ala Ala Phe Leu Ala Ser
            100                 105                 110

Pro Val Phe Asp Ala Val Asn Gly Phe Gly Gly Asn Gly Pro Tyr Val
        115                 120                 125

Asp Thr Ser Asn Phe Pro Val Thr Asn Val Pro Val Lys Ile Pro Asn
    130                 135                 140

Lys Thr Gly Gly Gly Cys Val Gln Asp Gly Ala Phe Val Asn Met Thr
145                 150                 155                 160

Val Thr Leu Gly Pro Gly Pro Ser Leu Glu Ser Asn Pro Arg Cys Leu
                165                 170                 175

Thr Arg Asp Phe Ser Tyr Trp Leu Ile Ser Arg Thr Leu Thr Lys Ala
            180                 185                 190
```

-continued

```
Val Val Asp Trp Thr Leu Glu Ala Ala Ser Phe Ala Val Phe Asp Phe
        195                 200                 205

Arg Leu Gln Gly Thr Gly Ile Glu Pro Glu Gly Met Thr Val His Ala
    210                 215                 220

Gly Gly His Leu Gly Val Gly Gly Asp Ile Gly Glu Ile Gly Asn Met
225                 230                 235                 240

Tyr Ser Ser Pro Gly Asp Pro Leu Phe Tyr Leu His His Ala Asn Leu
                245                 250                 255

Asp Arg Leu Trp Asp Gln Trp Gln Arg Lys Lys Phe Ala Gln Arg Val
            260                 265                 270

Gln Asp Met Thr Gly Pro Asp Thr Met Trp Ala Tyr Pro Phe Asn Phe
        275                 280                 285

Phe Gly Asp Val Pro Tyr Thr Asn Ile Thr Leu Glu Thr Leu Leu Asp
    290                 295                 300

Phe Lys Gly Leu Leu Gly Ser Ser Ser Ala Asp Arg Tyr Val Lys Ile
305                 310                 315                 320

Lys Asp Val Met Asp Ser Gln Gly Pro Asn Leu Cys Val Phe Tyr Lys
                325                 330                 335


<210> SEQ ID NO 32
<211> LENGTH: 370
<212> TYPE: PRT
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 32

Gln Gly Pro Cys Ser Thr Ile Arg Gln Arg Arg Ala Trp His Thr Leu
1               5                   10                  15

Ser Asn Ser Glu Lys Arg Ala Tyr Leu Asp Ala Glu Val Cys Leu Leu
            20                  25                  30

Gly Lys Ala Pro Lys Phe Gly Phe Glu Gly Ala Lys Asn Arg Phe Glu
        35                  40                  45

Glu Leu Gln Ala Ala His Gln Val Gln Ala Tyr Leu Ile His Gly Val
    50                  55                  60

Gly Ala Phe Leu Pro Phe His Arg Tyr Leu Met His Ala His Glu Thr
65                  70                  75                  80

Leu Leu Arg Thr Glu Cys Gly Tyr Gln Gly Ala Gln Pro Tyr Trp Asp
                85                  90                  95

Glu Thr Arg Asp Ala Gly Arg Val Ser Glu Ser Glu Ile Leu Asp Pro
            100                 105                 110

Asp Thr Gly Phe Gly Gly Asp Gly Val Gly Glu Arg Gly Cys Ile Ala
        115                 120                 125

Asp Gly Pro Phe Ala Gly Tyr Ile Asn Ser Ile Gly Pro Gly Tyr Arg
    130                 135                 140

Ile Thr Asp Arg Cys Ile Thr Arg Phe Val Asn Asn Thr Arg Ser Leu
145                 150                 155                 160

Met Ala Ser Pro Arg Phe Thr Asp Arg Cys Gln Gly Met Asn Arg Tyr
                165                 170                 175
```

-continued

```
Val Asp Val Trp Pro Cys Leu Glu Gly Asn Pro His Asn Ser Gly His
            180                 185                 190

Gly Ala Ile Ser Gly Leu Met Met Asp Pro Ile Ala Ser Pro Gly Asp
        195                 200                 205

Pro Ile Phe Tyr Leu His His Thr Trp Leu Asp Lys Leu Trp Trp Glu
    210                 215                 220

Trp Gln Ala Met Asp Leu Pro Arg Arg Leu Thr Asp Ile Gly Gly Arg
225                 230                 235                 240

Asn Thr Gln Asp Gly Ser Glu Gly Phe Pro Gly Ala Pro Pro Asn Ser
                245                 250                 255

Asn Gly Pro Asn Gly Ala Lys Arg Arg Ser Pro Ala Asp Gly Pro Ile
            260                 265                 270

Leu Ile Phe Pro Gly Asp Gly Gly Asn Ile Asp Trp Asp Glu Ile Asp
        275                 280                 285

Trp Glu Lys Ile Gly Phe Pro Gly Leu Gly Gly Gly Asn Gly Gly Pro
    290                 295                 300

Ile Gln Leu Pro Pro Gly Val Asp Val Pro Pro Glu Ala Leu Glu Pro
305                 310                 315                 320

Pro Glu Asp Ala Glu Pro Gln Glu Pro Arg Gly Asp Pro Gly Asp Val
                325                 330                 335

Thr Thr Leu Asn His Val Leu Lys Met Phe Gly Leu Val Pro Asp Ala
            340                 345                 350

Leu Ile Arg Asp Val Met Asp Ile Ala Gly Gly Thr Leu Cys Tyr Glu
        355                 360                 365

Tyr Val
    370


<210> SEQ ID NO 33
<211> LENGTH: 330
<212> TYPE: PRT
<213> ORGANISM: Paracedosporium putredinis

<400> SEQUENCE: 33

His Thr His Thr Leu Val Pro His Met Gln Asp Ser Thr Pro Cys Ile
1               5                   10                  15

Asn Pro Ser Leu Arg Arg Pro Trp Gln Ile Leu Ser Asp Gly Glu Lys
            20                  25                  30

Arg Ser Tyr Leu Asp Ala Gln Leu Cys Val Met Arg Thr Pro Gln Thr
        35                  40                  45

Leu Gly Leu Pro Gly Ala Arg Thr Arg Phe Glu Glu Leu Ala Ala Thr
    50                  55                  60

His Gln Ile Gly Ala Arg Ala Ser His Ala Thr Gly Thr Phe Phe Pro
65                  70                  75                  80

Tyr His Arg Tyr Leu Leu His Ala His Glu Ser Leu Leu Lys Glu Cys
                85                  90                  95

Gly Tyr His Ala Gly Leu Pro Tyr Trp Asp Glu Thr Arg Glu Ala Gly
            100                 105                 110
```

-continued

```
Asn Phe Ile Lys Ser Thr Ile Phe Glu Ser Gly Leu Gly Phe Gly Gly
        115                 120                 125

Phe Gly Ser Asp Leu Lys Gly Cys Ile Glu Asp Gly Pro Phe Ala Asn
    130                 135                 140

Leu Thr Ser Thr Ile Gly Pro Gly Phe Ser Leu Asn Glu His Cys Ile
145                 150                 155                 160

Ser Arg Ala Leu Asn Glu Thr Ala Gly Leu Lys Ala Ala Arg Glu Glu
                165                 170                 175

Val Asp Lys Cys Leu Glu Ala Asn Asp Tyr Thr Glu Met Trp Arg Cys
            180                 185                 190

Ala Tyr Thr Thr Pro His Arg Gly Gly His Gly Gly Val Gly Gly Thr
        195                 200                 205

Met Gly Asp Ala Leu Ala Ser Pro Gly Asp Pro Val Phe Tyr Val His
    210                 215                 220

His Ala Trp Val Asp Lys Ile Trp Trp Asp Trp Gln Glu Ala Asp Leu
225                 230                 235                 240

Asp Asn Arg Met Tyr Ala Ile Gly Gly Pro Ser Phe Gln Ser Pro Asp
                245                 250                 255

Ile Gly Phe Pro Glu Val Pro Gly Asp Val Glu Glu Glu Glu Ala Asn
            260                 265                 270

Ile Phe Gly Lys Pro Ser Glu Ala Ile Arg Arg Leu Gln Glu Leu Trp
        275                 280                 285

Ser Ser Ser Asp Pro Ser Arg Glu Thr Thr Leu Glu His Asn Leu Thr
    290                 295                 300

Leu Leu Gly Ile Ile Pro Asp Ile Asn Ile Ser Lys Val Met Asp Thr
305                 310                 315                 320

Arg Gly Gly Tyr Leu Cys Tyr Glu Tyr Val
                325                 330


<210> SEQ ID NO 34
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 34

tccgtaggtg aacctgcgg                                                  19


<210> SEQ ID NO 35
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 35

cgctgcgttc ttcatcg                                                    17


<210> SEQ ID NO 36
<211> LENGTH: 15
<212> TYPE: DNA
```

-continued

<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer
<220> FEATURE:
<221> NAME/KEY: y
<222> LOCATION: (2)..(2)
<223> OTHER INFORMATION: wherein y is C or T
<220> FEATURE:
<221> NAME/KEY: y
<222> LOCATION: (7)..(7)
<223> OTHER INFORMATION: wherein y is C or T
<220> FEATURE:
<221> NAME/KEY: d
<222> LOCATION: (8)..(8)
<223> OTHER INFORMATION: wherein a is A or G or T
<220> FEATURE:
<221> NAME/KEY: n
<222> LOCATION: (14)..(14)
<223> OTHER INFORMATION: wherein n is A or G or T or C
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (14)..(14)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 36

aytgggydta aagng 15

<210> SEQ ID NO 37
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: primer
<220> FEATURE:
<221> NAME/KEY: n
<222> LOCATION: (4)..(4)
<223> OTHER INFORMATION: wherein n is C or T or G or A
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (4)..(4)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: v
<222> LOCATION: (5)..(5)
<223> OTHER INFORMATION: wherein v is A or C or G

<400> SEQUENCE: 37

tacnvgggta tctaatcc 18

<210> SEQ ID NO 38
<211> LENGTH: 391
<212> TYPE: PRT
<213> ORGANISM: Agaricus bisporus

<400> SEQUENCE: 38

Ser Asp Lys Lys Ser Leu Met Pro Leu Val Gly Ile Pro Gly Glu Ile
1 5 10 15

Lys Asn Arg Leu Asn Ile Leu Asp Phe Val Lys Asn Asp Lys Phe Phe
20 25 30

Thr Leu Tyr Val Arg Ala Leu Gln Val Leu Gln Ala Arg Asp Gln Ser
35 40 45

-continued

```
Asp Tyr Ser Ser Phe Phe Gln Leu Gly Gly Ile His Gly Leu Pro Tyr
    50                  55                  60

Thr Glu Trp Ala Lys Ala Gln Pro Gln Leu His Leu Tyr Lys Ala Asn
65                  70                  75                  80

Tyr Cys Thr His Gly Thr Val Leu Phe Pro Thr Trp His Arg Ala Tyr
                85                  90                  95

Glu Ser Thr Trp Glu Gln Thr Leu Trp Glu Ala Ala Gly Thr Val Ala
            100                 105                 110

Gln Arg Phe Thr Thr Ser Asp Gln Ala Glu Trp Ile Gln Ala Ala Lys
        115                 120                 125

Asp Leu Arg Gln Pro Phe Trp Asp Trp Gly Tyr Trp Pro Asn Asp Pro
    130                 135                 140

Asp Phe Ile Gly Leu Pro Asp Gln Val Ile Arg Asp Lys Gln Val Glu
145                 150                 155                 160

Ile Thr Asp Tyr Asn Gly Thr Lys Ile Glu Val Glu Asn Pro Ile Leu
                165                 170                 175

His Tyr Lys Phe His Pro Ile Glu Pro Thr Phe Glu Gly Asp Phe Ala
            180                 185                 190

Gln Trp Gln Thr Thr Met Arg Tyr Pro Asp Val Gln Lys Gln Glu Asn
        195                 200                 205

Ile Glu Gly Met Ile Ala Gly Ile Lys Ala Ala Ala Pro Gly Phe Arg
    210                 215                 220

Glu Trp Thr Phe Asn Met Leu Thr Lys Asn Tyr Thr Trp Glu Leu Phe
225                 230                 235                 240

Ser Asn His Gly Ala Val Val Gly Ala His Ala Asn Ser Leu Glu Met
                245                 250                 255

Val His Asn Thr Val His Phe Leu Ile Gly Arg Asp Pro Thr Leu Asp
            260                 265                 270

Pro Leu Val Pro Gly His Met Gly Ser Val Pro His Ala Ala Phe Asp
        275                 280                 285

Pro Ile Phe Trp Met His His Cys Asn Val Asp Arg Leu Leu Ala Leu
    290                 295                 300

Trp Gln Thr Met Asn Tyr Asp Val Tyr Val Ser Glu Gly Met Asn Arg
305                 310                 315                 320

Glu Ala Thr Met Gly Leu Ile Pro Gly Gln Val Leu Thr Glu Asp Ser
                325                 330                 335

Pro Leu Glu Pro Phe Tyr Thr Lys Asn Gln Asp Pro Trp Gln Ser Asp
            340                 345                 350

Asp Leu Glu Asp Trp Glu Thr Leu Gly Phe Ser Tyr Pro Asp Phe Asp
        355                 360                 365

Pro Val Lys Gly Lys Ser Lys Glu Glu Lys Ser Val Tyr Ile Asn Asp
    370                 375                 380

Trp Val His Lys His Tyr Gly
385                 390
```

<210> SEQ ID NO 39
<211> LENGTH: 339

-continued

<212> TYPE: PRT
<213> ORGANISM: Vitis vinifera

<400> SEQUENCE: 39

```
Ala Pro Ile Gln Ala Pro Asp Ile Ser Lys Cys Gly Thr Ala Thr Val
1               5                   10                  15

Pro Asp Gly Val Thr Pro Thr Asn Cys Cys Pro Pro Val Thr Thr Lys
            20                  25                  30

Ile Ile Asp Phe Gln Leu Pro Ser Ser Gly Ser Pro Met Arg Thr Arg
        35                  40                  45

Pro Ala Ala His Leu Val Ser Lys Glu Tyr Leu Ala Lys Tyr Lys Lys
    50                  55                  60

Ala Ile Glu Leu Gln Lys Ala Leu Pro Asp Asp Asp Pro Arg Ser Phe
65                  70                  75                  80

Lys Gln Gln Ala Asn Val His Cys Thr Tyr Cys Gln Gly Ala Tyr Asp
                85                  90                  95

Gln Val Gly Tyr Thr Asp Leu Glu Leu Gln Val His Ala Ser Trp Leu
            100                 105                 110

Phe Leu Pro Phe His Arg Tyr Tyr Leu Tyr Phe Asn Glu Arg Ile Leu
        115                 120                 125

Ala Lys Leu Ile Asp Asp Pro Thr Phe Ala Leu Pro Tyr Trp Ala Trp
    130                 135                 140

Asp Asn Pro Asp Gly Met Tyr Met Pro Thr Ile Tyr Ala Ser Ser Pro
145                 150                 155                 160

Ser Ser Leu Tyr Asp Glu Lys Arg Asn Ala Lys His Leu Pro Pro Thr
                165                 170                 175

Val Ile Asp Leu Asp Tyr Asp Gly Thr Glu Pro Thr Ile Pro Asp Asp
            180                 185                 190

Glu Leu Lys Thr Asp Asn Leu Ala Ile Met Tyr Lys Gln Ile Val Ser
        195                 200                 205

Gly Ala Thr Thr Pro Lys Leu Phe Leu Gly Tyr Pro Tyr Arg Ala Gly
    210                 215                 220

Asp Ala Ile Asp Pro Gly Ala Gly Thr Leu Glu His Ala Pro His Asn
225                 230                 235                 240

Ile Val His Lys Trp Thr Gly Leu Ala Asp Lys Pro Ser Glu Asp Met
                245                 250                 255

Gly Asn Phe Tyr Thr Ala Gly Arg Asp Pro Ile Phe Phe Gly His His
            260                 265                 270

Ala Asn Val Asp Arg Met Trp Asn Ile Trp Lys Thr Ile Gly Gly Lys
        275                 280                 285

Asn Arg Lys Asp Phe Thr Asp Thr Asp Trp Leu Asp Ala Thr Phe Val
    290                 295                 300

Phe Tyr Asp Glu Asn Lys Gln Leu Val Lys Val Lys Val Ser Asp Cys
305                 310                 315                 320

Val Asp Thr Ser Lys Leu Arg Tyr Gln Tyr Gln Asp Ile Pro Ile Pro
                325                 330                 335

Trp Leu Pro
```

-continued

<210> SEQ ID NO 40
<211> LENGTH: 408
<212> TYPE: PRT
<213> ORGANISM: Aspergillus oryzea

<400> SEQUENCE: 40

Met Val Ala Leu Gln Ala Leu Ser Leu Gly Leu Leu Ala Ser Gln Ala
1 5 10 15

Leu Ala Phe Pro Ala Ala Ser Gln Gln Ala Ala Thr Ala Thr Leu Pro
20 25 30

Thr Thr Ala Ser Ser Ser Thr Ala Val Ala Ser Ser Gln Leu Asp Gln
35 40 45

Leu Ala Asn Phe Ala Tyr Asn Val Thr Thr Asp Ser Val Ala Gly Gly
50 55 60

Ser Glu Ser Lys Arg Gly Gly Cys Thr Leu Gln Asn Leu Arg Val Arg
65 70 75 80

Arg Asp Trp Arg Ala Phe Ser Lys Thr Gln Lys Lys Asp Tyr Ile Asn
85 90 95

Ser Val Leu Cys Leu Gln Lys Leu Pro Ser Arg Thr Pro Ala His Leu
100 105 110

Ala Pro Gly Ala Arg Thr Arg Tyr Asp Asp Phe Val Ala Thr His Ile
115 120 125

Asn Gln Thr Gln Ile Ile His Tyr Thr Gly Thr Phe Leu Ala Trp His
130 135 140

Arg Tyr Phe Ile Tyr Glu Phe Glu Gln Ala Leu Arg Asp Glu Cys Ser
145 150 155 160

Tyr Thr Gly Asp Tyr Pro Tyr Trp Asn Trp Gly Ala Asp Ala Asp Asn
165 170 175

Met Glu Lys Ser Gln Val Phe Asp Gly Ser Glu Thr Ser Met Ser Gly
180 185 190

Asn Gly Glu Tyr Ile Pro Asn Gln Gly Asp Ile Lys Leu Leu Leu Gly
195 200 205

Asn Tyr Pro Ala Ile Asp Leu Pro Pro Gly Ser Gly Gly Gly Cys Val
210 215 220

Thr Ser Gly Pro Phe Lys Asp Tyr Lys Leu Asn Leu Gly Pro Ala Ala
225 230 235 240

Leu Ser Leu Pro Gly Gly Asn Met Thr Ala Ala Ala Asn Pro Leu Thr
245 250 255

Tyr Asn Pro Arg Cys Met Lys Arg Ser Leu Thr Thr Glu Ile Leu Gln
260 265 270

Arg Tyr Asn Thr Phe Pro Lys Ile Val Glu Leu Ile Leu Asp Ser Asp
275 280 285

Asp Ile Trp Asp Phe Gln Met Thr Met Gln Gly Val Pro Gly Ser Gly
290 295 300

Ser Ile Gly Val His Gly Gly Gly His Tyr Ser Met Gly Gly Asp Pro
305 310 315 320

Gly Arg Asp Val Tyr Val Ser Pro Gly Asp Thr Ala Phe Trp Leu His
325 330 335

-continued

```
His Gly Met Ile Asp Arg Val Trp Trp Ile Trp Gln Asn Leu Asp Leu
            340                 345                 350

Arg Lys Arg Gln Asn Ala Ile Ser Gly Thr Gly Thr Phe Met Asn Asn
        355                 360                 365

Pro Ala Ser Pro Asn Thr Thr Leu Asp Thr Val Ile Asp Leu Gly Tyr
    370                 375                 380

Ala Asn Gly Gly Pro Ile Ala Met Arg Asp Leu Met Ser Thr Thr Ala
385                 390                 395                 400

Gly Pro Phe Cys Tyr Val Tyr Leu
                405
```

<210> SEQ ID NO 41
<211> LENGTH: 281
<212> TYPE: PRT
<213> ORGANISM: Streptomyces castaneoglobisporus

<400> SEQUENCE: 41

```
Met Thr Val Arg Lys Asn Gln Ala Thr Leu Thr Ala Asp Glu Lys Arg
1               5                   10                  15

Arg Phe Val Ala Ala Val Leu Glu Leu Lys Arg Ser Gly Arg Tyr Asp
            20                  25                  30

Glu Phe Val Arg Thr His Asn Glu Phe Ile Met Ser Asp Thr Asp Ser
        35                  40                  45

Gly Glu Arg Thr Gly His Arg Ser Pro Ser Phe Leu Pro Trp His Arg
    50                  55                  60

Arg Phe Leu Leu Asp Phe Glu Gln Ala Leu Gln Ser Val Asp Ser Ser
65                  70                  75                  80

Val Thr Leu Pro Tyr Trp Asp Trp Ser Ala Asp Arg Thr Val Arg Ala
                85                  90                  95

Ser Leu Trp Ala Pro Asp Phe Leu Gly Gly Thr Gly Arg Ser Thr Asp
            100                 105                 110

Gly Arg Val Met Asp Gly Pro Phe Ala Ala Ser Thr Gly Asn Trp Pro
        115                 120                 125

Ile Asn Val Arg Val Asp Ser Arg Thr Tyr Leu Arg Arg Ser Leu Gly
    130                 135                 140

Gly Ser Val Ala Glu Leu Pro Thr Arg Ala Glu Val Glu Ser Val Leu
145                 150                 155                 160

Ala Ile Ser Ala Tyr Asp Leu Pro Pro Tyr Asn Ser Ala Ser Glu Gly
                165                 170                 175

Phe Arg Asn His Leu Glu Gly Trp Arg Gly Val Asn Leu His Asn Arg
            180                 185                 190

Val His Val Trp Val Gly Gly Gln Met Ala Thr Gly Val Ser Pro Asn
        195                 200                 205

Asp Pro Val Phe Trp Leu His His Ala Tyr Val Asp Lys Leu Trp Ala
    210                 215                 220

Glu Trp Gln Arg Arg His Pro Asp Ser Ala Tyr Val Pro Thr Gly Gly
225                 230                 235                 240
```

-continued

```
Thr Pro Asp Val Val Asp Leu Asn Glu Thr Met Lys Pro Trp Asn Thr
                245                 250                 255

Val Arg Pro Ala Asp Leu Leu Asp His Thr Ala Tyr Tyr Thr Phe Asp
            260                 265                 270

Ala Leu Glu His His His His His His
        275                 280


<210> SEQ ID NO 42
<211> LENGTH: 303
<212> TYPE: PRT
<213> ORGANISM: Bacillus megaterium

<400> SEQUENCE: 42

Met Gly Asn Lys Tyr Arg Val Arg Lys Asn Val Leu His Leu Thr Asp
1               5                   10                  15

Thr Glu Lys Arg Asp Phe Val Arg Thr Val Leu Ile Leu Lys Glu Lys
            20                  25                  30

Gly Ile Tyr Asp Arg Tyr Ile Ala Trp His Gly Ala Ala Gly Lys Phe
        35                  40                  45

His Thr Pro Pro Gly Ser Asp Arg Asn Ala Ala His Met Ser Ser Ala
    50                  55                  60

Phe Leu Pro Trp His Arg Glu Tyr Leu Leu Arg Phe Glu Arg Asp Leu
65                  70                  75                  80

Gln Ser Ile Asn Pro Glu Val Thr Leu Pro Tyr Trp Glu Trp Glu Thr
                85                  90                  95

Asp Ala Gln Met Gln Asp Pro Ser Gln Ser Gln Ile Trp Ser Ala Asp
            100                 105                 110

Phe Met Gly Gly Asn Gly Asn Pro Ile Lys Asp Phe Ile Val Asp Thr
        115                 120                 125

Gly Pro Phe Ala Ala Gly Arg Trp Thr Thr Ile Asp Glu Gln Gly Asn
    130                 135                 140

Pro Ser Gly Gly Leu Lys Arg Asn Phe Gly Ala Thr Lys Glu Ala Pro
145                 150                 155                 160

Thr Leu Pro Thr Arg Asp Asp Val Leu Asn Ala Leu Lys Ile Thr Gln
                165                 170                 175

Tyr Asp Thr Pro Pro Trp Asp Met Thr Ser Gln Asn Ser Phe Arg Asn
            180                 185                 190

Gln Leu Glu Gly Phe Ile Asn Gly Pro Gln Leu His Asp Arg Val His
        195                 200                 205

Arg Trp Val Gly Gly Gln Met Gly Val Val Pro Thr Ala Pro Asn Asp
    210                 215                 220

Pro Val Phe Phe Leu His His Ala Asn Val Asp Arg Ile Trp Ala Val
225                 230                 235                 240

Trp Gln Ile Ile His Arg Asn Gln Asn Tyr Gln Pro Met Lys Asn Gly
                245                 250                 255

Pro Phe Gly Gln Asn Phe Arg Asp Pro Met Tyr Pro Trp Asn Thr Thr
            260                 265                 270
```

-continued

```
Pro Glu Asp Val Met Asn His Arg Lys Leu Gly Tyr Val Tyr Asp Ile
        275                 280                 285

Glu Leu Arg Lys Ser Lys Arg Ser Ser His His His His His His
    290                 295                 300


<210> SEQ ID NO 43
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: copper binding site
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (2)..(8)
<223> OTHER INFORMATION: wherein X is any amino acid
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (10)..(17)
<223> OTHER INFORMATION: wherein X is any amino acid

<400> SEQUENCE: 43

His Xaa Xaa Xaa Xaa Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1               5                   10                  15

Xaa His

<210> SEQ ID NO 44
<211> LENGTH: 28
<212> TYPE: PRT
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: copper binding site
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (2)..(4)
<223> OTHER INFORMATION: wherein X is any amino acid
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (6)..(27)
<223> OTHER INFORMATION: wherein X is any amino acid

<400> SEQUENCE: 44

His Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1               5                   10                  15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His
            20                  25


<210> SEQ ID NO 45
<211> LENGTH: 29
<212> TYPE: PRT
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: copper binding site
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (2)..(4)
<223> OTHER INFORMATION: wherein X is any amino acid
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (6)..(28)
<223> OTHER INFORMATION: wherein X is any amino acid

<400> SEQUENCE: 45
```

-continued

```
His Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1               5                   10                  15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His
            20                  25


<210> SEQ ID NO 46
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: copper binding site
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (2)..(4)
<223> OTHER INFORMATION: wherein X is any amino acid
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (6)..(29)
<223> OTHER INFORMATION: wherein X is any amino acid

<400> SEQUENCE: 46

His Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1               5                   10                  15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His
            20                  25                  30


<210> SEQ ID NO 47
<211> LENGTH: 31
<212> TYPE: PRT
<213> ORGANISM: artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: copper binding site
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (2)..(4)
<223> OTHER INFORMATION: wherein X is any amino acid
<220> FEATURE:
<221> NAME/KEY: X
<222> LOCATION: (6)..(30)
<223> OTHER INFORMATION: wherein X is any amino acid

<400> SEQUENCE: 47

His Xaa Xaa Xaa His Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1               5                   10                  15

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa His
            20                  25                  30
```

1. An isolated nucleic acid molecule encoding a β-etherase polypeptide wherein said polypeptide comprises copper and further wherein the activity of said polypeptide is independent of NAD^+ and/or glutathione.

2. The isolated nucleic acid molecule according to claim **1**, wherein said isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:

i) a nucleotide sequence as set forth in SEQ ID NO: 18, SEQ ID NO:17. SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, or SEQ ID NO: 25;

ii) a nucleotide sequence wherein said sequence is degenerate as a result of the genetic code to the nucleotide sequence defined in (i);

iii) a nucleic acid molecule comprising a nucleotide sequence the complementary strand of which hybridizes under stringent hybridisation conditions to sequence set forth in SEQ ID NO: 18, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO:24 or SEQ ID NO: 25;

iv) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:

26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32 or SEQ ID NO: 33;

v) a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence wherein said amino acid sequence is modified by addition, deletion or substitution of at least one amino acid residue as represented in iv) above and has β-etherase activity.

3-9. (canceled)

10. An isolated β-etherase polypeptide wherein said polypeptide comprises copper and further wherein the activity of said polypeptide is independent of NAD^+ and/or glutathione.

11. The isolated polypeptide according to claim **10**, wherein said isolated polypeptide is selected from the group consisting of:

i) a polypeptide comprising or consisting of an amino acid sequence as set forth in SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31. SEQ ID NO: 32 OR SEQ ID NO: 33;

ii) a modified polypeptide comprising or consisting of a modified amino acid sequence wherein said polypeptide is modified by addition, deletion or substitution of at least one amino acid residue of the sequence set forth in SEQ ID NO: 26. SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32 or SEQ ID NO: 33, and which has β-etherase activity.

12-18. (canceled)

19. A vector comprising the nucleic acid molecule according to claim **1**.

20. The vector according to claim **19**. wherein the vector is an expression vector adapted for expression in a heterologous microbial host cell.

21. A cell transformed or transfected with the nucleic acid molecule according to claim **1**.

22. The cell according to claim **21**, wherein said cell is a heterologous host cell wherein said heterologous host cell does not naturally express the nucleic acid molecule.

23. The cell according to claim **21**, wherein said cell is a bacterial cell, a fungal cell or a yeast cell.

24. (canceled)

25. The cell according to claim **23**, wherein said fungal cell is an Aspergillus sp. cell. or wherein said fungal cell is not a *Parascedosporium* sp cell.

26. (canceled)

27. A composition comprising one or more polypeptides according to claim **10**.

28. A composition according to claim **27**, wherein said composition comprises at least the polypeptide set forth in SEQ ID NO: 9 or 26.

29. A composition according to claim **27**, wherein said one more polypeptides are set forth in SEQ ID NO: 26, 27, 28, 29, 30, 31, 32 and 33.

30. A composition according to claim **27** wherein said composition further comprises one or more polypeptides for the saccharification of lignocellulose selected from the group consisting of cellulases, lytic polysaccharide monooxygenases, carbohydrate esterases, hemicellulases, glycosylhydrolases, endoglucanases, cellobiohydrolases, beta-glucosidases, xylanases, mannases, cellobiose dehydrogenases, and beta-xylosidases.

31. A method for the modification of plant biomass comprising the following steps:

i) contacting plant biomass with the composition according to

claim **27** to form a reaction mixture; and ii) incubating said reaction mixture under conditions which cleave β-ether linkages present in the plant biomass to obtain depolymerised lignin units.

32. The method according to claim **31**, wherein;

said method comprises a further step of extracting said depolymerised lignin units from the reaction mixture;

said method comprises a further step of contacting said reaction mixture with a composition comprising one or more polypeptides for the saccharification of the processed lignocellulose; and/or

said method comprises extracting di- and/or monosaccharides.

33. The method according to claim **31**, wherein:

said depolymerised lignin units are selected from the group consisting of flavones and p-coumaric acid; and/or

said plant biomass is wheat straw or sugarcane bagasse.

34. The method according to claim **33** wherein said flavones are tricin.

35-36. (canceled)

37. The method according to claim **32**, wherein said saccharification composition comprises or consist of one or more polypeptides selected from the group consisting of cellulases, lytic polysaccharide monooxygenases, carbohydrate esterases, hemicellulases, glycosylhydrolases, endoglucanases, cellobiohydrolases, beta-glucosidases, xylanases, mannases, cellobiose dehydrogenases, and beta-xylosidases.

38. (canceled)

39. A method for the manufacture of a β-etherase polypeptide comprising the following steps:

i) providing the cell according to claim **21** and cell culture medium,

ii) culturing the cell in i) above to express a β-eherase polypeptide wherein said polypeptide comprises copper and further wherein the activity of said polypeptide is independent of NAD^+ and/or glutathione; and optionally,

iii) isolating said polypeptide from the cell or cell culture medium.

40. The method according to claim **39** wherein said polypeptide is isolated under denaturing conditions.

* * * * *