1 **Identification of a β -glucosidase from the *Mucor circinelloides* genome by Peptide**

2 **Pattern Recognition**

3 Yuhong Huang^{a,b,c}, Peter Kamp Busk^a, Morten Nedergaard Grell^a, Hai Zhao^b, Lene Lange^{a*}

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

¹**Abbreviations:** PPR: Peptide pattern recognition; GH family: glycoside hydrolase family;
pNPG: p-nitrophenyl- β -D-glucopyranoside; MUG: 4-methylumbelliferyl β -D-
glucopyranoside

Abstract: *Mucor circinelloides* produces plant cell wall degrading enzymes that allow it to grow on complex polysaccharides. Although the genome of *M. circinelloides* has been sequenced, only few plant cell wall degrading enzymes are annotated in this species. We applied peptide pattern recognition, which is a non-alignment based method for sequence analysis to map conserved sequences in glycoside hydrolase families. The conserved sequences were used to identify similar genes in the *M. circinelloides* genome. We found 12 different novel genes encoding members of the GH3, GH5, GH9, GH16, GH38, GH47 and GH125 families in *M. circinelloides*. One of the two GH3-encoding genes was predicted to encode a β -glucosidase (EC 3.2.1.21). We expressed this gene in *Pichia pastoris* KM71H and found that the purified recombinant protein had relative high β -glucosidase activity (1.73 U/mg) at pH5 and 50 °C. The K_m and V_{max} with p-nitrophenyl- β -D-glucopyranoside as substrate was 0.20 mM and 2.41 U/mg, respectively. The enzyme was not inhibited by glucose and retained 84 % activity at glucose concentrations up to 140 mM. Although zygomycetes are not considered to be important degraders of lignocellulosic biomass in nature, the present finding of an active β -glucosidase in *M. circinelloides* demonstrates that enzymes from this group of fungi have a potential for cellulose degradation.

Keywords: peptide pattern recognition; *Mucor circinelloides*; β -glucosidase; glycoside hydrolase 3

1. Introduction

Lignocellulose from plant cell walls is the most abundant organic complex in agricultural waste [1] and decomposition of the recalcitrant cell wall biomass in agricultural crop residues and waste presents an important scientific challenge [2]. Although the polysaccharides can be broken down by carbohydrate-active enzymes, the extreme variety of monosaccharide structures and sugar cross-linkages in the cell wall requires careful selection of the appropriate enzymes for each material [3]. Thus, there is a need to discover more efficient and cost effective enzymes for depolymerization of agricultural crop residues and waste streams to valuable products [4]. Presently, the number of carbohydrate-active enzymes increases dramatically due to the wealth of new information from genomic, transcriptomic, and secretomic projects. The clustered sequences are annotated based on similarity search, such as batch blast in the NCBI protein database [5], the UniProtKB protein database [6], the microorganism database [7] or other databases. The protein sequences for carbohydrate catabolism can be further analyzed through the use of the CAZy database and with different bioinformatic tools. The most used method for sequence comparison is alignment, through which regions with similar sequences in two or more biological polymers can be found. However, current alignment methods are limited to handling novel proteins sequences with high similarity and it is difficult to analyze large numbers of highly divergent proteins [8].

In nature, agricultural crop residue degrading microorganisms include bacteria, yeasts and filamentous fungi. Especially filamentous fungi belonging to *Ascomycota* and *Basidiomycota* [9,10,11,12] are generally considered the most potent cellulase and hemicellulase producers. However, zygomycetous fungi, such as *Rhizopus oryzae* [6,13,14],

Rhizomucor miehei [15,16], *Thermomucor indicae-seudaticae* [17] and *M. circinelloides* [18], also play an important role in degradation of agricultural crop residue. It was reported that *M. circinelloides* can produce a complete set of cellulose degrading enzymes [19], indicating that this fungus has a potential for being used in biomass conversion. However, although the genome of *M. circinelloides* is available, only few plant cell wall degrading enzymes are annotated for this species.

A novel method, peptide pattern recognition (PPR) [20], was applied to mine the *M. circinelloides* genome for genes encoding glycoside hydrolases. PPR is a method for non-alignment based sequence analysis, which can simultaneously compare multiple even highly variable sequences and find the characteristic features. It is a potent approach for revealing new protein subfamily grouping, with a shared specific function, even for proteins with very little sequence identity [20]. By this approach, 12 different novel genes encoding GH3, GH5, GH9, GH16, GH38, GH47 and GH125 family members were discovered in the *M. circinelloides* genome and subsequently confirmed by searching GenBank at the National Center for Biotechnology Information (NCBI) for homologues using the BLAST tool [21].

Of the two GH3 encoding genes found, one was predicted by PPR to encode a β -glucosidase (EC 3.2.1.21). An important role of this enzyme is to prevent product inhibition of different cellulases such as endoglucanase and cellobiohydrolase by removal of the cellobiose produced by these enzymes [22]. Furthermore, most of the currently used commercial cellulase blends generally exhibit rather low β -glucosidase activity [23]. High level β -glucosidase production is crucial to further develop the sustainable industrial application of cellulose. Therefore, the discovered β -glucosidase gene was heterologously

87 expressed in *Pichia pastoris* and the recombinant β -glucosidase purified and characterized

88 biochemically.

89

2. Materials and methods

2.1. PPR data mining of *M. circinelloides* genome to find interesting enzyme functions

Glycoside hydrolase families including cellulose, mannan and xylan degrading enzymes were identified in the CAZy database (<http://www.cazy.org/>) (Table A.1, Appendix data). The full length protein sequences of these enzymes were found and the resulting subgroups related to function as described by Busk and Lange [20]. Finally, the n-mer peptides for each subgroup were used to find genes encoding related enzymes in the *M. circinelloides* genome as described by Busk and Lange [20].

Full length genes were obtained by BLAST search (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) against the *M. circinelloides* reference genome (<http://genome.jgi-psf.org/Mucci1/Mucci1.home.html>). Only genes for which a complete open reading frame could be identified were further analyzed.

2.2. β -glucosidase structure and phylogeny

The 3D structure of the predicted *M. circinelloides* β -glucosidase protein was modeled by Swiss model (<http://swissmodel.expasy.org/>).

Compute pI/Mw (http://web.expasy.org/compute_pi/) was applied to predict the theoretical isoelectric point and molecular weight.

N-GlycoSite (<http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html>) was used to predict the potential N-glycosylation sites.

The signal peptide was predicted by SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>).

The phylogenetic analysis was performed by the following steps: (1) the *M. circinelloides* β -glucosidase protein sequence was used as query to search the NCBI non-redundant protein sequence database for homologous, full length sequences using BLAST; (2) the protein sequences of β -glucosidases from different microorganisms were downloaded from NCBI; (3) the *M. circinelloides* β -glucosidase and all downloaded protein sequences were aligned using ClustalX [24]; and (4) a phylogenetic tree was generated using neighbor-joining algorithm in MEGA ver5.0 [25].

2.3. Cloning and heterologous expression of the predicted β -glucosidase in *Pichia pastoris*

The cDNA sequence of the *M. circinelloides* β -glucosidase was synthesized and cloned into pUC57 by GenScript (USA). The cDNA was excised from pUC57 and inserted between the *EcoRI* and *XbaI* sites of pPICZ α B (Invitrogen, EasySelect *Pichia* Expression Kit). The resulting plasmid was digested with *PmeI* and transformed into *P. pastoris* KM71H by electroporation according to the EasySelect manual. Multi-copy recombinants were selected by passing single clones onto YPDS plates with increasing concentration of Zeocin (100, 500, 1000 and 2000 μ g/ml). Expression of recombinant β -glucosidase was performed according to the EasySelect manual.

2.4. Enzyme assay and total protein determination

The *P. pastoris* expression culture was centrifuged, and aliquots of the supernatant used to assay β -glucosidase activity according to Krisch et al. [26] with some modification. The reaction was carried out using 5 mM p-nitrophenyl- β -D-glucopyranoside (pNPG) (Sigma) as the substrate in McIlvaine Buffer (pH 5) [27]. A 15 μ l portion of the supernatant, or 15 μ l of

appropriately diluted enzyme, was mixed with 150 μ l substrate and incubated at 50 °C for 30 min. 30 μ l of the reaction sample was transferred to a microtiter plate and terminated by adding 50 μ l 1 M Na₂CO₃. The developed color was read at 405 nm by a Packard Spectra Count plate reader. One unit of β -glucosidase activity was defined as the amount of enzyme required to produce 1 μ mol *p*-nitrophenol (pNP) per minute under the above assay conditions.

Total protein in the supernatant was determined by the Bradford method using the BCA Protein Assay Kit (Thermo Scientific) and bovine serum albumin (BSA) as standard.

2.5. Purification of expressed β -glucosidase

Scale-up expression was performed according to the EasySelect manual (Invitrogen). Cells were incubated at 30 °C, 250 rpm for 5 days. The supernatant was harvested by centrifugation at 1500 x g for 5 minutes at room temperature and filtered through a 0.2 μ m filter (Minisart Syringe Filters).

The His-tagged β -glucosidase was purified by fast protein liquid chromatography (FPLC) (ÄKTA Purifier) using the UNICORN method and a 1 ml HisTrap FF crude affinity column (GE Healthcare). First, the column was equilibrated with binding buffer (20 mM sodium phosphate, 500 mM NaCl, 30 mM imidazole, pH 7.4) with a flow rate of 1 ml/min. Then, 100 ml sample was loaded onto the column, followed by washing with binding buffer until the absorbance reached a steady baseline. The His-tagged β -glucosidase was finally eluted with elution buffer (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4).

2.6. Characterization of purified β -glucosidase

2.6.1. SDS-PAGE electrophoresis and zymography

Samples of 40 µl crude or purified enzyme were mixed with 10 µl loading buffer, incubated at 70 °C for 10 min and loaded on a 10% (w/v) polyacrylamide gel. After electrophoresis, gels were silver stained using the Pierce Silver Stain Kit (Thermo scientific) as instructed by the manufacturer.

The zymography assay was performed as described by Liu et al. [28] using 0.5 mM 4-methylumbelliferyl β-D-glucopyranoside (MUG) (Sigma) as substrate.

2.6.2. Effects of temperature and pH on β-glucosidase activity

To investigate the effect of temperature on the purified enzyme, reactions were carried out at different temperatures (30, 40, 50, 60 or 70 °C) but otherwise as described in section 2.4. In order to study thermo stability, the purified enzyme was pre-incubated for 60 min at 30, 40, 50, 60 or 70 °C in McIlvaine buffer (pH 5), and then residual enzyme activity was measured.

To determine the pH optimum of the purified enzyme, reactions were carried out over a pH range of 3-8 at otherwise standard conditions. To study pH stability, the purified enzyme was incubated in McIlvaine buffer of different pH (4, 5, 6, 7 or 8) at 4 °C for 1 h. Residual enzyme activity was then determined.

The protein concentration of purified recombinant enzyme was calculated according to the molar extinction coefficient of the *M. circinelloides* β-glucosidase protein sequence (<http://encorbio.com/protocols/Prot-MW-Abs.htm>). The absorbance of the protein at the ultraviolet wavelength of 280 nm was measured by Nanodrop1000 (Thermo Scientific).

2.6.3. Kinetic analysis

The Michaelis-Menten kinetics of the purified β -glucosidase was determined by pNPG concentrations ranging from 0.1 to 5 mM in McIlvaine buffer (pH 5) and by cellobiose concentrations ranging from 1 to 15 mM in McIlvaine buffer (pH 5). The enzyme was diluted and β -glucosidase activity was measured using pNPG as substrate as describe above (section 2.4), except that incubation time was only 10 min. The kinetic analysis using different concentrations of cellobiose in McIlvaine buffer (pH 5) was performed as follows: 15 μ l enzyme and 150 μ l substrate was incubated at 50 °C for 10 min. The reaction was terminated by incubating at 95 °C for 5 min. The glucose concentration was measured using D-glucose-HK kit (Megazyme) as instructed by the manufacturer. One unit of β -glucosidase activity was defined as the amount of enzyme required to produce 1 μ mol of glucose per minute under the above assay conditions. The Michaelis-Menten constants K_m and V_{max} were determined from Hanes-Wolf plots [29].

2.6.4. Glucose tolerance of β -glucosidase

To investigate the glucose tolerance of the purified β -glucosidase, reactions were carried out at glucose concentrations ranging from 0 to 140 mM. The residual activity (glucose tolerance) was measured as described above (section 2.4).

3. Results and discussion

3.1. Mining the *M. circinelloides* genome with PPR to find interesting enzymes

The PPR tool was used to predict genes encoding cellulose, mannan and xylan degrading enzymes in the *M. circinelloides* genome. The result included 12 different novel genes encoding enzymes belonging to the GH3, GH5, GH9, GH16, GH38, GH47 and GH125 families. Additionally, a gene was identified encoding a GH45 with 92 % identity to an endo-beta-D-1,4-glucanase from *M. circinelloides* (GenBank: BAD95808.1) (Table 1). Information on the location of novel genes in the genome of *M. circinelloides* CBS 277.49 (v1.0) can be found in Table A.2 (Appendix data). The protein and nucleotide sequences of the genes in Table 1 are available in appendix data (Predicted protein sequence and Predicted nucleotide sequence, respectively).

One of the GH3 members was predicted to be a β -glucosidase. This result is different from the situation for the closely related *R. oryzae*, in which none of the GH3 family proteins are β -glucosidases [13]. Two GH5 members from the *M. circinelloides* genome have putative exo- β -1,3-glucanase activity. Furthermore, a putative GH9 protein (endoglucanase) was identified although most GH9 proteins are from bacteria, plants and animals [30]. Six GH16 proteins were found in the *M. circinelloides* genome. These enzymes efficiently degrade β -1,3-1,4-glucans. However, neither cellobiohydrolases (GH6 and GH7) nor xylanases (GH10 and GH11) or GH61 proteins were found in the *M. circinelloides* genome. This result is similar to what was reported for the *R. oryzae* genome [13].

3.2. Bioinformatics analysis of the predicted β -glucosidase gene and protein

The size of the predicted β -glucosidase gene is 2492 bp including five introns located at positions 182-252, 520-579, 1307-1365, 1435-1490 and 1738-1796. The deduced protein consists of 729 amino acid residues with a theoretical isoelectric point of 5.61 and a molecular weight of 78.24 kDa, predicted by Compute pI/Mw. Six potential N-glycosylation sites were predicted at residues 94, 230, 293, 396, 487 and 627 by N-GlycoSite. A signal peptide with a cleavage site between amino acid 25 and 26 was predicted by SignalP 4.1. Conserved domain prediction suggests that the β -glucosidase belongs to GH3 carrying a GH3 N-terminal domain (Pfam00933, located at residues 38-318) and a GH3 C-terminal domain (Pfam01915, located at residues 362-613). Analysis of the predicted amino acid sequence revealed 57% identity with β -glucosidase from *R. miehei* (GenBank: CAP58431.2) [16]. Comparing with other related β -glucosidases, a conserved catalytic domain (LLKEELGFKGFVQSDWG, located at residues 244-260) was found in the predicted amino acid sequence. Liu et al. [28] and Jeya et al. [31] also reported a conserved GFVMSD and GFVMTD sequence, respectively. These residues were identified as part of the catalytic domain of GH3 family proteins. The catalytic aspartic acid residue [32] of the predicted β -glucosidase from *M. circinelloides* is located at residue 258.

A phylogenetic tree was constructed to compare the β -glucosidase from *M. circinelloides* to β -glucosidases from other fungi (Fig. 1). The β -glucosidase from *M. circinelloides* is most closely related to *R. miehei* β -glucosidase. The β -glucosidases from the Mucorales form a group that is separate from the β -glucosidases from the Basidiomycota and Ascomycota,

Fungal β -glucosidases are abundant. However, at present, only two GH1 β -glucosidase crystal structures from *Phanerochaete chrysosporium* and *Trichoderma reesei* and two GH3

β -glucosidase crystal structures from *Trichoderma reesei* and *Aspergillus aculeatus* are available at NCBI. Homology modeling studies show that the domains of the *M. circinelloides* β -glucosidase are similar to that of the GH3 β -glucosidase from *Trichoderma reesei* (Fig. 2). Although the sequence identity is only 46%, it is obvious that the residues important for substrate binding and catalysis are conserved.

3.3. Heterologous expression of the predicted β -glucosidase in *Pichia pastoris*

The recombinant β -glucosidase was expressed in *P. pastoris* and purified after five days incubation when the activity was highest (0.81 U/ml) (Fig. A.1, Appendix data). Even though the purified β -glucosidase had lower specific activity (1.73 U/mg) on pNPG than recombinant fungal β -glucosidase (GH3) from *Aspergillus fumigatus* [28], *Rhizomucor miehei* [16], *Paecilomyces thermophila* [33], *Myceliophthora thermophila* [34], *Periconia sp.* [35] and *Chrysosporium lucknowense* [36], this predicted β -glucosidase still has the potential for cellulose degradation as much higher specificity for the hydrolysis of pNPG and glucose tolerance we found as follows.

3.4. Characterization of the purified recombinant β -glucosidase

3.4.1. SDS-PAGE and zymography analysis of the purified recombinant β -glucosidase

The purified recombinant β -glucosidase migrated as a single band on the SDS-PAGE gel (Fig. 3) with an apparent molecular weight of approximately 75 kDa, which is close to the predicted molecular weight of 78.2 kDa. Zymography analysis indicated that the β -glucosidase can hydrolyze MUG and create a clear fluorescent band in the gel after being

visualized under UV light at 365 nm (Fig. 3).

3.4.2. Effects of temperature and pH on the activity of the predicted β -glucosidase

The highest activity of the β -glucosidase from *M. circinelloides* was achieved at 50 °C (Fig. 4), although the enzyme lost 42% of its activity after one hour incubation at 40 °C in McIlvaine buffer (pH 5). The pH optimum of the β -glucosidase was determined to be 5 (Fig. 4). Krisch et al. [26] and Dashtban and Qin [35] reported similar results for β -glucosidases from *R. miehei* and *Periconia sp.*, respectively. The pH stability results indicated that the β -glucosidase can tolerate pH values from 4 to 8 with the high residual activity at pH 7 (Fig. 4).

3.4.3. Kinetic analysis of the predicted β -glucosidase

The K_m (Fig. A.2, Appendix data) of the *M. circinelloides* β -glucosidase was calculated to be 0.20 mM when using pNPG as substrate, which is lower than the reported Novozymes S188 (Sigma) and β -glucosidase from *Aspergillus niger*, BGL1 from *Trichoderma reesei* [37,38,39], indicating that the recombinant β -glucosidase from *M. circinelloides* has much higher specificity for the hydrolysis of pNPG. However, the V_{max} and K_{eff} ($=V_{max}/K_m$) of the *M. circinelloides* β -glucosidase were 2.41 U/mg and 12.05 U/mg/mM, respectively (Fig. A.2, Appendix data), which are much lower than reported for *R. miehei* [26]. As pNPG and cellobiose have different structures, so the kinetic analysis was also performed on cellobiose, the K_m and V_{max} were 4.55 mM and 2.48 U/mg, respectively. But the results still indicate that the recombinant β -glucosidase from *M. circinelloides* is degrading pNPG fast.

3.4.4. Glucose tolerance of the predicted β -glucosidase

Glucose tolerance was investigated using pNPG as substrate at different glucose concentrations. The β -glucosidase activity was unaffected by glucose concentrations below 100 mM (Fig. 5). When the glucose concentration increased above 100 mM, the β -glucosidase activity was slightly inhibited. However, the enzyme remained greater than 84% active at 140 mM glucose. It was reported that β -glucosidase was very sensitive to glucose inhibition and most of the microbial β -glucosidases were completely inhibited by glucose [34,40,41]. The interesting phenomenon that this enzyme was found to be insensitive to glucose is similar to what was shown for a β -glucosidase from *Thermoanaerobacterium thermosaccharolyticum* [42]. It was reported that β -glucosidases were crucial for complete hydrolysis of cellulose as they can hydrolyze the cellobiose and small cello-oligosaccharides into glucose. In this study, the recombinant β -glucosidase displays glucose tolerance and is highly active at glucose concentration up to 140 mM, indicating that this enzyme not only can tolerate the glucose inhibition, but also can decrease the concentration of the cellobiose, which further synergistically optimize the actions of endoglucanases and exoglucanases to completely and efficiently in cellulose depolymerization [43]. Thus, the high glucose tolerance makes the purified recombinant β -glucosidase of great potential and interest for biotechnological applications.

4. Conclusion

This study is the first report on the prediction, expression, purification and characterization of a β -glucosidase from *M. circinelloides*. In this work, the PPR method was applied to mine the *M. circinelloides* genome for genes encoding glycoside hydrolases. Twelve different novel genes encoding members of the GH3, GH5, GH9, GH16, GH38, GH47 and GH125 families were found in the genome. Of the two GH3 genes, the one predicted to encode β -glucosidase was expressed in *Pichia pastoris*. The specific activity of the purified recombinant β -glucosidase was 1.73 U/mg when incubated at pH 5 and 50 °C for 30 min. The $V_{\max}$ (2.41 U/mg) was found with pNPG as substrate. The enzyme displays glucose tolerance and is highly active at glucose concentration up to 140 mM. This work shows that a zygomycete species like *M. circinelloides* possess highly active carbohydrate-modifying enzymes of industrial relevance.

322 **Acknowledgements**

323 We thank Dr. Torres Martinez and Joint Genome Institute for providing the *Mucor*
324 *circinelloides* genome data. This study was supported by the Sino-Danish Center (SDC), The
325 Danish Strategic Research Council projects 2101-08-0041 (BioRef), 09-065251 (Bio4Bio)
326 and Novozymes A/S. The work conducted by JGI, U.S. Department of Energy Joint Genome
327 Institute is supported by the Office of Science of the U.S. Department of Energy under
328 Contract No. DE-AC02-05CH11231.

329

References

- [1] Sweeney MD, Xu F. Biomass Converting Enzymes as Industrial Biocatalysts for Fuels and Chemicals: Recent Developments. *Catalysts* 2012; 2: 244-63.
- [2] Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 2007; 315: 804-7.
- [3] Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 2009; 37: 5.
- [4] King BC, Waxman KD, Nenni NV, Walker LP, Bergstrom GC, Gibson DM. Arsenal of plant cell wall degrading enzymes reflects host preference among plant pathogenic fungi. *Biotechnol Biofuels* 2011; 4: 1754-6834.
- [5] Tartar A, Wheeler M, Zhou X, Coy M, Boucias D, Scharf M. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol Biofuels* 2009; 2: 25.
- [6] Couturier M, Navarro D, Olive C, Chevret D, Haon M, Favel A, et al. Post-genomic analyses of fungal lignocellulosic biomass degradation reveal the unexpected potential of the plant pathogen *Ustilago maydis*. *BMC Genomics* 2012; 13: 1471-2164.
- [7] Sun J, Tian C, Diamond S, Glass NL. Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in *Neurospora crassa*. *Eukaryot Cell* 2012; 11: 482-93.

352 [8] Huang W, Umbach DM, Li L. Accurate anchoring alignment of divergent sequences.
353 Bioinformatics 2006; 22: 29-34.

354 [9] De Souza WR, De Gouvea PF, Savoldi M, Malavazi I, De Souza Bernardes LA,
355 Goldman MH, et al. Transcriptome analysis of *Aspergillus niger* grown on sugarcane
356 bagasse. Biotechnol Biofuels 2011; 4: 1754-6834.

357 [10] Hakkinen M, Arvas M, Oja M, Aro N, Penttila M, Saloheimo M, et al. Re-annotation
358 of the CAZy genes of *Trichoderma reesei* and transcription in the presence of
359 lignocellulosic substrates. Microb Cell Fact 2012; 11: 1475-2859.

360 [11] Le Crom S, Schackwitz W, Pennacchio L, Magnuson JK, Culley DE, Collett JR, et al.
361 Tracking the roots of cellulase hyperproduction by the fungus *Trichoderma reesei*
362 using massively parallel DNA sequencing. Proc Natl Acad Sci USA 2009; 106:
363 16151-6.

364 [12] Vitikainen M, Arvas M, Pakula T, Oja M, Penttila M, Saloheimo M. Array
365 comparative genomic hybridization analysis of *Trichoderma reesei* strains with
366 enhanced cellulase production properties. BMC Genomics 2010; 11: 1471-2164.

367 [13] Battaglia E, Benoit I, van den Brink J, Wiebenga A, Coutinho PM, Henrissat B, et al.
368 Carbohydrate-active enzymes from the zygomycete fungus *Rhizopus oryzae*: a highly
369 specialized approach to carbohydrate degradation depicted at genome level. BMC
370 Genomics 2011; 12: 1471-2164.

371 [14] Ghosh B, Ray RR. Current commercial perspective of *Rhizopus oryzae*: A review. J
372 Applied Sci 2011; 11: 2470-86.

373 [15] Fawzi EM. Highly thermostable xylanase purified from *Rhizomucor miehei* NRL 3169.
374 Acta Biol Hung 2011; 62: 85-94.

375 [16] Tako M, Toth A, L GN, Krisch J, Vagvolgyi C, Papp T. A new beta-glucosidase gene
376 from the zygomycete fungus *Rhizomucor miehei*. Antonie Van Leeuwenhoek 2010; 97:
377 1-10.

378 [17] Busk P, Lange L. Cellulolytic potential of thermophilic species from four fungal
379 orders. AMB Express 2013; 3: 47.

380 [18] Baba Y, Shimonaka A, Koga J, Kubota H, Kono T. Alternative splicing produces two
381 endoglucanases with one or two carbohydrate-binding modules in *Mucor*
382 *circinelloides*. J Bacteriol 2005; 187: 3045-51.

383 [19] Saha BC. Production, purification and properties of endoglucanase from a newly
384 isolated strain of *Mucor circinelloides*. Process Biochem 2004; 39: 1871-6.

385 [20] Busk PK, Lange L. Function-based classification of carbohydrate-active enzymes by
386 recognition of short, conserved peptide motifs. Appl Environ Microbiol 2013; 79:
387 3380-91.

388 [21] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search
389 tool. J Mol Biol 1990; 215.

390 [22] Zhang YH, Lynd LR. Toward an aggregated understanding of enzymatic hydrolysis of
391 cellulose: noncomplexed cellulase systems. Biotechnol Bioeng 2004; 88: 797-824.

392 [23] Chen M, Zhao J, Xia L. Enzymatic hydrolysis of maize straw polysaccharides for the
393 production of reducing sugars. Carbohydr Polym 2008; 71: 411-5.

394 [24] Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, G. HD. The
395 ClustalX windows interface: flexible strategies for multiple sequence alignment aided
396 by quality analysis tools. *Nucleic Acids Res* 1997; 25: 4876-82.

397 [25] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular
398 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
399 maximum parsimony methods. *Mol Biol Evol* 2011; 28: 2731-9.

400 [26] Krisch J, Bencsik O, Papp T, Vagvolgyi C, Tako M. Characterization of a beta-
401 glucosidase with transgalactosylation capacity from the zygomycete *Rhizomucor*
402 *miehei*. *Bioresource Technol* 2012; 114: 555-60.

403 [27] McIlvaine TC. A Buffer Solution for Colorimetric comparison. *J Biol Chem* 1921; 49:
404 183-6.

405 [28] Liu D, Zhang R, Yang X, Zhang Z, Song S, Miao Y, et al. Characterization of a
406 thermostable beta-glucosidase from *Aspergillus fumigatus* Z5, and its functional
407 expression in *Pichia pastoris* X33. *Microb Cell Fact* 2012; 11: 1475-2859.

408 [29] Haldane JBS. Graphical Methods in Enzyme Chemistry. *Nature* 1957; 179: 832-.

409 [30] Ni J, Takehara M, Watanabe H. Identification of activity related amino acid mutations
410 of a GH9 termite cellulase. *Bioresource Technol* 2010; 101: 6438-43.

411 [31] Jeya M, Joo AR, Lee KM, Tiwari MK, Kim SH, Lee JK. Characterization of beta-
412 glucosidase from a strain of *Penicillium purpurogenum* KJS506. *Appl Microbiol Biot*
413 2010; 86: 1473-84.

- 414 [32] Wong WK, Ali A, Chan WK, Ho V, Lee NT. The cloning, expression and
415 characterization of a cellobiase gene encoding a secretory enzyme from *Cellulomonas*
416 *biazotea*. *Gene* 1998; 207: 79-86.
- 417 [33] Yan Q, Hua C, Yang S, Li Y, Jiang Z. High level expression of extracellular secretion
418 of a beta-glucosidase gene (PtBglu3) from *Paecilomyces thermophila* in *Pichia*
419 *pastoris*. *Protein Expr Purif* 2012; 84: 64-72.
- 420 [34] Karnaouri A, Topakas E, Paschos T, Taouki I, Christakopoulos P. Cloning, expression
421 and characterization of an ethanol tolerant GH3 beta-glucosidase from
422 *Myceliophthora thermophila*. *PeerJ* 2013; 1: e46.
- 423 [35] Dashtban M, Qin W. Overexpression of an exotic thermotolerant beta-glucosidase in
424 *trichoderma reesei* and its significant increase in cellulolytic activity and
425 saccharification of barley straw. *Microb Cell Fact* 2012; 11: 1475-2859.
- 426 [36] Dotsenko GS, Sinitsyna OA, Hinz SWA, Wery J, Sinitsyn AP. Characterization of a
427 GH family 3 β -glycoside hydrolase from *Chrysosporium lucknowense* and its
428 application to the hydrolysis of β -glucan and xylan. *Bioresource Technol* 2012; 112:
429 345-9.
- 430 [37] Chauve M, Mathis H, Huc D, Casanave D, Monot F, Lopes Ferreira N. Comparative
431 kinetic analysis of two fungal beta-glucosidases. *Biotechnol Biofuels* 2010; 3: 1754-
432 6834.
- 433 [38] Himmel ME, Adney WS, Fox JW, Mitchell DJ, Baker JO. Isolation and
434 characterization of two forms of beta-D-glucosidase from *Aspergillus niger*. *Appl*
435 *Biochem Biotech* 1993; 40: 213-25.

- [39] Seidle HF, Marten I, Shoseyov O, Huber RE. Physical and kinetic properties of the family 3 beta-glucosidase from *Aspergillus niger* which is important for cellulose breakdown. *Protein J* 2004; 23: 11-23.
- [40] Karnchanatat A, Petsom A, Sangvanich P, Piaphukiew J, Whalley AJS, Reynolds CD, et al. Purification and biochemical characterization of an extracellular β -glucosidase from the wood-decaying fungus *Daldinia eschscholzii* (Ehrenb.:Fr.) Rehm. *FEMS Microbiol Lett* 2007; 270: 162-70.
- [41] Parry NJ, Beever DE, Owen E, Vandenberghe I, Van Beeumen J, Bhat MK. Biochemical characterization and mechanism of action of a thermostable beta-glucosidase purified from *Thermoascus aurantiacus*. *Biochem J* 2001; 353: 117-27.
- [42] Pei J, Pang Q, Zhao L, Fan S, Shi H. *Thermoanaerobacterium thermosaccharolyticum* beta-glucosidase: a glucose-tolerant enzyme with high specific activity for cellobiose. *Biotechnol Biofuels* 2012; 5: 1754-6834.
- [43] Lima MA, Oliveira-Neto M, Kadowaki MA, Rosseto FR, Prates ET, Squina FM, et al. *Aspergillus niger* beta-glucosidase has a cellulase-like tadpole molecular shape: insights into glycoside hydrolase family 3 (GH3) beta-glucosidase structure and function. *J Biol Chem* 2013; 288: 32991-3005.

1 **Table 1** Enzyme discoveries with Peptide pattern recognition (PPR) technology in the *Mucor*
2 *circinelloides* genome

3

Gene	EC	Known enzyme activity	GH	score	hits	new gene (yes/no)
>4013 6	3.2.1.52	beta-N-acetylhexosaminidase	3	6,28	8	yes
>3802 2	3.2.1.21	beta-glucosidase	3	3,55	6	yes
>48 1 rev	3.2.1.58	exo-beta-1,3-glucanase	5	13,58	24	yes
>234 2	3.2.1.45	beta-1,3-glucanase	5	3,22	9	yes
>2961 7 rev	3.2.1.4	endo-beta-1,4-glucanase	9	5,17	8	yes
>1569 5	3.2.1.39	beta 1,3-glucanase	16	9,22	17	yes
>2910 6	3.2.1.6	1,3(4)-beta-glucanase	16	6,83	14	yes
>4543 5	2.4.1.207	cell wall glucanosyltransferase	16	6,7	12	yes
>153 8 rev	3.2.1.24	AMS1-alpha-mannosidase	38	11,99	17	yes
>2433 2	3.2.1.4	endo-beta-D-1,4-glucanase	45	33,34	66	no
>3030 8	3.2.1.113	alpha-mannosidase	47	15,92	28	yes
>3956 2		alpha-1,2-mannosidase	125	9,6	18	yes

4
5

Figure captions

Fig. 1. Phylogenetic tree based on the amino acid sequences of the predicted *Mucor circinelloides* GH3 and related protein sequences obtained from NCBI. The phylogenetic tree was reconstructed by the neighbor-joining method using CLUSTAL X and MEGA ver. 5.0. The levels of bootstrap support are indicated at the nodes. The scale bar represents 0.1 amino acid substitutions per position.

Fig. 2. Homology model of the predicted *Mucor circinelloides* β -glucosidase (A). Crystal structure of a glycoside hydrolase family 3 beta-glucosidase, Bgl1 from *Hypocrea jecorina* (<http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?uid=105724>) (B).

Fig. 3. SDS-PAGE and zymography analysis of the recombinant *Mucor circinelloides* β -glucosidase. Lane M: PageRuler Prestained Protein Ladder, 10 – 170 kDa (Thermo scientific); lane 1: crude protein samples from *Pichia pastoris* KM71H/pPICZ α B/ β -glucosidase; lane 2: His-tag purified protein samples from *Pichia pastoris* KM71H/ pPICZ α B/ β -glucosidase; lane 3: β -glucosidase MUG zymogram.

Fig. 4. Effects of temperature (up) and pH (down) on the activity and stability of the purified recombinant *Mucor circinelloides* β -glucosidase. Values are averages counted from analyses; error bars represent standard deviation.

Fig. 5. Effect of glucose on the activity of the purified recombinant *Mucor circinelloides*

23 β -glucosidase. Values are averages counted from analyses; error bars represent standard
24 deviation.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5