Fungal glucuronoyl esterases: genome mining based enzyme discovery and biochemical characterization

Adiphol Dilokpimol[a,1], Miia R. Mäkelä[b,1], Gabriella Cerullo[c,1], Miaomiao Zhou[a,g], Simona Varriale[c], Loknath Gidijala[d], Joana L.A. Brás[e], Peter Jütten[f], Alexander Piechot[f], Raymond Verhaert[d], Vincenza Faraco[c,*], Kristiina S. Hilden[b,*], Ronald P. de Vries[a,*]

[a]Fungal Physiology, Westerdijk Fungal Biodiversity Institute & Fungal Molecular Physiology, Utrecht University, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

[b]Division of Microbiology and Biotechnology, Department of Food and Environmental Sciences, University of Helsinki, Helsinki, Finland

[c]Department of Chemical Sciences, University of Naples "Federico II", Complesso Universitario Monte S. Angelo Via Cintia, 4 IT-80126, Naples, Italy

[d]ProteoNic B.V., J.H. Oortweg 19-21, 2333 CH Leiden, The Netherlands

[e]NZYTech B.V., Estrada do Paço do Lumiar, Campus do Lumiar, Edifício E – 1º andar, 1649-038 Lisboa, Portugal

[f]Taros Chemicals GmbH & Co. KG, Emil-Figge-Str. 76a, 44227 Dortmund, Germany

[g]Current address: Academie voor de Technologie van Gezondheid en Milieu, Avans Hogeschool, Lovensdijkstraat 61, 4818 AJ Breda, The Netherlands

[*]Corresponding authors.

E-mail addresses: r.devries@westerdijkinstitute.nl (R.P. de Vries), kristiina.s.hilden@helsinki.fi (K.S. Hilden); vincenza.faraco@unina.it (V. Faraco).

[1]Equal contribution.

Abstract

4-*O*-methyl-D-glucuronic acid (MeGlcA) is a side-residue of glucuronoarabinoxylan and can form ester linkages to lignin, contributing significantly to the strength and rigidity of the plant cell wall. Glucuronoyl esterases (4-*O*-methyl-glucuronoyl methylesterases, GEs) can cleave this ester bond, and therefore may play a significant role as an auxiliary enzyme in biomass saccharification for the production of biofuels and biochemicals. GEs belong to a relatively new family of carbohydrate esterases (CE15) in the CAZy database (www.cazy.org), and so far around ten fungal GEs have been characterized. To explore additional GE enzymes, we used a genome mining strategy. BLAST analysis with characterized GEs against approximately 250 publicly accessible fungal genomes identified more than 150 putative fungal GEs, which were classified into eight phylogenetic sub-groups. To validate the genome mining strategy, 21 selected GEs from both ascomycete and basidiomycete fungi were heterologously produced in *Pichia pastoris*. Of these enzymes, 18 were active against benzyl D-glucuronate demonstrating the suitability of our genome mining strategy for enzyme discovery.

Highlights

- Over 150 putative fungal GEs were identified by genome mining
- Putative fungal GEs were classified into 8 phylogenetic sub-groups
- New ascomycete and basidiomycete fungal GEs were produced in *Pichia pastoris*
- Several of the GEs showed very high activity
- New GEs showed potential in plant biomass processing applications

Keywords (4-6 words)

Glucuronoyl esterase, glucuronic acid, genome mining, plant cell wall, fungi

Abbreviations

CE, carbohydrate esterase; GE, glucuronoyl esterase.

Introduction

4-*O*-methyl-D-glucuronic acid (MeGlcA) is a side-residue of xylan (β-1,4-linked D-xylose) that is found in both glucuronoxylan and glucuronoarabinoxylan, which are the principle components present in the secondary cell walls of eudicotyledonous plants and both cell wall layers of commelinoid monocots, respectively (Fig. 1) [1-3]. A large proportion of MeGlcA in xylan can form ester linkages to lignin alcohol; for example 30% and 40% of MeGlcA are esterified to lignin in beechwood and birchwood, respectively [4, 5]. In nature, these lignin-carbohydrate complexes (LCCs) contribute significantly to the strength and rigidity of the plant cell wall rendering it recalcitrant to digestion. However, they impede the industrial applications of plant biomass by restricting the removal of lignin e.g. from cellulosic pulp in pulping processes and hindering efficient enzymatic hydrolysis of biomass in bioethanol production [6-9].

Glucuronoyl esterases (4-*O*-methyl-glucuronoyl methylesterases, GEs) can cleave the ester bond between MeGlcA and lignin, and therefore may play a significant role as auxiliary enzymes in biomass saccharification for the production of biofuels and biochemicals. The first GE was reported in 2006 from a white-rot like fungus *Schizophyllum commune* [10], and belongs to carbohydrate esterase family 15 (CE15) in the CAZy database [11, 12]. From 182 members in CE15, only 21 are from fungi, and of these so far only around 10 GEs have been characterized (Table 1). Among these, the structures of the *Trichoderma reesei* (*Hypocrea jecorina*) Cip2 [13] and the *Myceliophthora thermophila* (*Sporotrichum thermophile*) StGE2 [14] have been resolved by X-ray crystallography. The first structure revealed the Ser-His-Glu as the putative catalytic triad of GEs, whereas in the latter case the catalytic serine mutant in complex with methyl 4-*O*-methyl-β-D-glucopyranuronate was also reported revealing substrate binding within the active site and indicating possible catalytic mechanism of GEs.

In 2014, a European Union (EU) collaborative project 'Optimized esterase biocatalysts for cost-effective industrial production' (OPTIBIOCAT, www.optibiocat.eu) was granted by 7th Framework Programme (FP7), which aims to replace chemical processes by enzymatic bioconversion via transesterification of esterases such as GEs for the production of cosmetics. To explore additional fungal GE enzymes, we used a genome mining analysis towards approximately 250 publicly accessible fungal genomes [15]. In this study, we report the genome mining strategy to identify novel fungal GEs and verify the strategy by biochemical characterization of the heterologously produced selected GEs, from both ascomycete and basidiomycete fungi, representing different phylogenetic sub-groups.

Materials and methods

Bioinformatics

Genome mining was performed by BLASTP search against 247 published fungal genomes [15] using 15 amino acid sequences from characterized and putative GEs (A.1, A.2 in Supplementary materials). All resulting amino acid sequences with an expected value lower than $1E^{-40}$ were collected. Duplicate, unusually long and incomplete sequences as well as sequences with ambiguous amino acids (X) were discarded. Signal peptides were predicted using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/ [16]) and removed from all candidate sequences. The sequences were aligned using Multiple Alignment using Fast Fourier Transform (MAFFT) [17]. Phylogenetic analysis was performed using the maximal likelihood method with complete deletion of gaps and the Poisson correction distance of substitution rates (statistical support for phylogenetic grouping was estimated by 1000 bootstrap re-samplings) of the Molecular Evolutionary Genetics Analysis (MEGA 7) program [18]. A few feruloyl esterase sequences were included as an outgroup. Theoretical molecular masses and p*I* values were calculated by ExPASy–ProtParam tool (http://www.expasy.ch/tools/protparam.html [19]).

Cloning of ge *genes*

The genes encoding the selected GEs without signal peptide were codon optimized and synthesized for expression in *P. pastoris* by NZYTech (Lisbon, Portugal). The gene products were digested by *EcoR*I *and Not*I (Thermo Fisher Scientific), and cloned in frame with α-factor secretion signal in pPNic706 (ProteoNic, Leiden, the Netherlands). The obtained plasmids were purified from *Escherichia coli* DH5α (Invitrogen) transformants selected on Luria Bertani medium supplemented with 50 µg/mL kanamycin, fully sequenced (Macrogen, Amsterdam, the Netherlands), linearised by *Sal*I (Thermo Fisher Scientific), and transformed into *P. pastoris* strain GS115 *his*4 according to the manufacturer's recommendation.

Ten transformants were selected for the enzyme production screening, which was performed in 96 deep-well plates containing 0.8 mL medium. The selected clones were grown first in buffered minimal glycerol medium (1% yeast nitrogen base, 0.1 M potassium phosphate buffer pH 6.5, and

1 % w/v glycerol). The plates were sealed with AeraSeal™ (Sigma Aldrich) and were incubated overnight at 30°C, 900 rpm (INFORS HT Microtron, Bottmingen, Switzerland). A volume of cells equal to an OD_{600} of 1.0 was harvested and resuspended in 0.8 mL buffered minimal methanol medium (1% yeast nitrogen base, 0.1 M potassium phosphate buffer pH 6.5, and 0.5 % methanol) for induction. The induction was done at 30°C, 900 rpm for 72 h before being harvested. The cultures were supplemented with 80 µL of 0.5% (v/v) methanol every 24 h.

Production and biochemical properties of recombinant GEs

P. pastoris transformants were grown according to [20]. Induction was continued for 96 h at 28°C with methanol being supplemented to 0.5% (v/v) every 24 h. Culture supernatants were harvested (4000 × g, 4°C, 20 min), filtered (0.22 µm; Merck Millipore, Darmstadt, Germany) or concentrated (10 kDa cut off; Merck Millipore) and stored at -20°C prior further analysis. Molecular mass determination and deglycosylation were performed as previously described [20]. Protein concentrations were assessed from SDS-PAGE gels by densitometry method using ImageJ program [21] with bovine serum albumin (Pierce, Thermo Scientific) as standard.

Enzyme activity assay of GEs

Activity of the recombinant GEs towards benzyl D-glucuronate (Taros Chemicals, Dortmund, Germany) was performed in 200 µL reaction mixtures adapted from [22]. The reactions were performed in the presence of 2 mM substrate, 73 mM phosphate buffer, pH 6.0, and 50 µL of culture supernatant at 45°C for 30 min. Detection of glucuronic acid release was performed by using D-Glucuronic/D-Galacturonic Acid Assay Kit (Megazyme, Wicklow, Ireland) according to the manufacturer's recommendation. The culture supernatant of *P. pastoris* harboring pPNic706 plasmid without insert was used as negative control. All assays were performed in triplicate. One unit was defined as the amount of enzyme de-esterifying 1 μmol of benzyl-D-glucuronic acid ester per min under the assay conditions.

Results and discussion

Genome mining and phylogenetic analysis of novel fungal GEs

To identify the putative fungal GEs, a genome mining strategy was conducted by BLAST analysis with characterized and putative GEs against the published fungal genomes [15]. More than 150 putative fungal GEs were identified, which can be classified into 8 phylogenetic sub-groups (Fig. 2, A.1, A.2 in Supplementary materials). The first characterized GE (*S. commune*, ScGE, [10, 23]) located to Sub-group 4 together with GEs from the white-rot fungi *Phanerochaete carnosa* (PcGCE, [24]), *Phanerochaete chrysosporium* (PcGE1, PcGE2; [25]) and *Cerrena unicolor* (CuGE, [26]) (Fig. 2). The ascomycete GEs, *Trichoderma reesei* GE (Cip2, [12]) and *Podospora anserina* GE (PaGE1, [27]) clustered in Sub-group 5. Sub-group 8 consisted of the GEs from the ascomycete fungi *Myceliophthora thermophila* (StGE1, [28]) and *Neurospora crassa* (NcGE, [29]), whereas Sub-group 1 consisted of a second GE from *M. thermophila* (StGE2, [30]). No characterized GE belongs to Sub-group 2, 3, 6 and 7. Sub-group 3 consists of more than 30 members and Sub-group 7 consists of 15 members, whereas Sub-group 2 and 6 are small sub-groups containing 3-4 GE candidates. One GE candidate from an anaerobic fungus *Piromyces* sp. E2 (PirGE1) did not locate to any of the sub-groups. This ungrouped sequence may develop into a new sub-group if homologs for it are discovered. Two characterized bacterial GEs were included in the analysis, and clustered separately from the fungal GEs (Fig. 2).

Recently, a new classification of GEs was reported based on peptide pattern recognition (PPR) [31], which separate putative GEs into 24 PPR groups. Fungal GEs were clustered in PPR groups 1, 8, and 18. In comparison to our phylogenetic classification, the members from PPR group 8 belonged to phylogenetic Sub-group 1, whereas the members from PPR group 1 were divided in different phylogenetic groups. PPR group 18 contained only one member (GenBank XP_001832002.2 from *Coprinopsis cinerea*) representing an unusually long sequence (containing 3438 amino acids), hence it was not included in our phylogenetic analysis. In addition, a new database for Carboxylic Ester Hydrolases (CEH) was launched - CASTLE (CArboxylic eSTer hydrolase, http://castle.cbe.iastate.edu/, Iowa State University [32]). However, GEs are currently grouped together with acetyl xylan esterases in CEH8 in CASTLE database.

Sequence analysis and catalytic triad of selected fungal GEs

Twenty-one candidates (five characterized and 16 putative fungal GEs) were selected from both ascomycete and basidiomycete fungi with focus on wood rotting fungi (e.g. *Dichomitus squalens*, *P. carnosa*, *Schizophyllum commune*, *Stereum hirsutum*), saprophytic fungi living on dead plant or herbivore dung (e.g. *Podospora anserina*, *Ascobolus immersus*, *Apiospora montagnei*), plant pathogens (e.g. *Botryosphaeria dothidea*, *Stagonospora nodorum*, *Leptosphaeria maculans*) as well as industrially exploited fungi (e.g. *T. reesei*, *Penicillium rubens*), covering all eight sub-groups from the phylogenetic tree, including one ungrouped sequence (PirGE1), for heterologous production using *P. pastoris* as a host and subsequent biochemical characterization (Table 2). The selection of the number of putative GEs was solely based on the size of the phylogenetic sub-group. The amino acid sequence alignment of 16 putative fungal GEs and all characterized GEs are present in A.3 in Supplementary materials. The fungal GEs were relatively conserved and the signature motif of CE15 family (G-C-S-R-X-G, [30]) was well aligned, except for PirGE1 which has Tyr instead of Arg. In addition, two bacterial GEs (CesA and MZ0003) have Val and His, respectively, instead of Cys in the signature motif. Among the catalytic triad, Ser and His are well conserved in all sequences, whereas Glu is not very conserved among CE15 enzymes and is substituted by Asp, Gln, Asn and Ala, as well as Ser (StGE2, PaGE2 – *Podospora anserina*, DsGE1 – *Dichomitus squalens*) and Cys (LmGE1 – *Leptosphaeria maculans*, CesA and MZ0003) [33].

Biochemical properties of selected fungal GEs

The putative GE-encoding genes were heterologously expressed in *P. pastoris*. Only two (PaGE2 and AbGE1 – *Agaricus bisporus*) out of 21 GE candidates were not successfully produced. The production level varied from 2-336 mg/L, and four enzymes (HsGE1 – *Hypholoma sublateritium*, PcGCE, Cip2, ShGE1 – *Stereum hirsutum*) were produced up to 300 mg/L. From the 19 produced GE candidates, 18 were active towards benzyl-D-glucuronic acid ester (Table 2). The highest specific activity (>1,000 nkat/mg) was detected for PcGCE, HsGE1 and DsGE1. SnGE1 – *Stagonospora nodorum* and LmGE1 showed low activity (0.156 nkat/ml and 0.097 nkat/ml, respectively) and were produced at low level as they were not visible in Coomassie blue stained SDS-PAGE gel. CcGE1 protein from *Coprinopsis cinerea* was highly produced but not active towards the tested substrate on different pH (4-8).

Conclusions

In the present study, we showed that genome mining is a powerful strategy for enzyme discovery to identify fungal GE encoding genes. Our phylogenetical analysis categorized the putative fungal

GEs into eight sub-groups. We further demonstrated that from 16 putative fungal GEs, 13 possessed GE activity towards benzyl D-glucuronate. The members from Sub-groups 1, 4, 5 and 8 were previously characterized and shown to possess GE activity (Table 1). Here we demonstrated that the candidates from Sub-groups 3, 6, and 7 also possessed GE activity (Table 2). Because of the limited availability of substrates used for the assessment of GE activity, currently it is not possible to verify if the phylogenetical grouping also reflects functional differences among GEs, such as substrate specificity or possible site of action.

In comparison with the previously characterized fungal GEs used in this study, most of the new GEs showed comparable activity. This indicates that they may have potential in saccharification of plant biomass or other industrial applications.

Funding This work was supported by the European Union, Grant agreement no: 613868 (OPTIBIOCAT).

Acknowledgments

The authors would like to thank Prof. Paul Christakopoulos (Luleå University of Technology, Sweden) for providing the glucuronoyl esterase assay method.

Appendix A. Supplementary data

A.1 Phylogenetic tree of the (putative) fungal GEs. △, characterized GEs. The sequences used for BLASTP search in genome mining analysis are indicated as ◇ for characterized GEs and ○ for putative GEs. Filled symbols indicate selected GEs for characterization in the present study. Bact.

indicates group of bacterial GEs. U indicates ungrouped sequences. Feruloyl esterases (FAEs) were used as an outgroup. The same symbols are used in Fig. 2.

A.2 Sequences of characterized and putative GEs used in the present study. Protein families (Pfam) and carbohydrate binding motifs (CBM) were predicted using MOTIF Search tool from GenomeNet (http://www.genome.jp/tools/motif/, Bioinformatics Center - Kyoto University). Selected GEs for characterization in the present study are marked in light blue.

A.3 Amino acid sequence alignment of characterized and selected GEs (full names of the enzymes are given in Tables 1 and 2). The sequence alignment was performed using Multiple Alignment using Fast Fourier Transform (MAFFT) [17] and visualized using Easy Sequencing in Postscript (ESPript - http://espript.ibcp.fr, [34]). The structures of *Hypocrea jecorina* Cip2 (3PIC, [13]) were used as a secondary structure depiction. ▼ indicates catalytic residues.

References

[1] York WS, O'Neill MA. Biochemical control of xylan biosynthesis — which end is up? Curr Opin Plant Biol 2008;11:258-265.

[2] Scheller HV, Ulvskov P. Hemicelluloses. Annu Rev Plant Biol 2010;61:263-289.

[3] Ebringerová A. Structural diversity and application potential of hemicelluloses. Macromol Symp 2005;232:1-12.

[4] Takahashi N, Koshijima T. Molecular properties of lignin-carbohydrate complexes from beech (*Fagus crenata*) and pine (*Pinus densiflora*) woods. Wood Sci and Technol 1988;22:177-189.

[5] Watanabe T, Koshijima T. Evidence for an ester linkage between lignin and glucuronic acid in lignin–carbohydrate complexes by DDQ-Oxidation. Agric Biol Chem 1988;52:2953-2955.

[6] Koshijima T, Watanabe T. *Association between lignin and carbohydrates in wood and other plant tissues*, Springer Berlin Heidelberg 2003.

[7] Várnai A, Siika-aho M, Viikari L. Restriction of the enzymatic hydrolysis of steam-pretreated spruce by lignin and hemicellulose. Enzyme Microb Technol 2010;46:185-193.

[8] Himmel ME, Ding S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, *et al.* Biomass recalcitrance: Engineering plants and enzymes for biofuels production. Science 2007;315:804-807.

[9] Min D-y, Yang C, Chiang V, Jameel H, Chang H-m. The influence of lignin–carbohydrate complexes on the cellulase-mediated saccharification II: Transgenic hybrid poplars (*Populus nigra* L. and *Populus maximowiczii* A.). Fuel 2014;116:56-62.

[10] Špániková S, Biely P. Glucuronoyl esterase – Novel carbohydrate esterase produced by *Schizophyllum commune*. FEBS Lett 2006;580:4597-4601.

[11] Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 2014;42:D490-D495.

[12] Li X-L, Špániková S, de Vries RP, Biely P. Identification of genes encoding microbial glucuronoyl esterases. FEBS Lett 2007;581:4029-4035.

[13] Pokkuluri PR, Duke NEC, Wood SJ, Cotta MA, Li X-L, Biely P, *et al.* Structure of the catalytic domain of glucuronoyl esterase Cip2 from *Hypocrea jecorina*. Proteins: Struct, Funct, Bioinf 2011;79:2588-2592.

[14] Charavgi MD, Dimarogona M, Topakas E, Christakopoulos P, Chrysina ED. The structure of a novel glucuronoyl esterase from *Myceliophthora thermophila* gives new insights into its role as a potential biocatalyst. Acta Crystallogr Sect D-Biol Crystallogr 2013;69:63-73.

[15] Dilokpimol A, Mäkelä MR, Aguilar-Pontes MV, Benoit-Gelber I, Hildén KS, de Vries RP. Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. Biotechnol Biofuels 2016;9:231.

[16] Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth 2011;8:785-786.

[17] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772-780.

[18] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 2016;33:1887.

[19] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al.* Protein identification and analysis tools on the ExPASy Server. In: Walker JM (editor). The proteomics protocols handbook, Totowa, NJ: Humana Press; 2005, p. 571-607.

[20] Dilokpimol A, Mäkelä MR, Mansouri S, Belova O, Waterstraat M, Bunzel M, *et al.* Expanding the feruloyl esterase gene family of *Aspergillus niger* by characterization of a feruloyl esterase, FaeC. N Biotechnol 2017;37, Part B:200-209.

[21] Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Meth 2012;9:671-675.

[22] Sunner H, Charavgi M-D, Olsson L, Topakas E, Christakopoulos P. Glucuronoyl esterase screening and characterization assays utilizing commercially available benzyl glucuronic acid ester. Molecules 2015;20:17807.

[23] Wong DWS, Chan VJ, McCormack AA, Hirsch J, Biely P. Functional cloning and expression of the *Schizophyllum commune* glucuronoyl esterase gene and characterization of the recombinant enzyme. Biotechnol Res Int 2012;2012:951267.

[24] Tsai AYL, Canam T, Gorzsás A, Mellerowicz EJ, Campbell MM, Master ER. Constitutive expression of a fungal glucuronoyl esterase in *Arabidopsis* reveals altered cell wall composition and structure. Plant Biotech J 2012;10:1077-1087.

[25] Duranová M, Hirsch J, Kolenová K, Biely P. Fungal glucuronoyl esterases and substrate uronic acid recognition. Biosci Biotechnol Biochem 2009;73:2483-2487.

[26] d'Errico C, Jørgensen JO, Krogh KBRM, Spodsberg N, Madsen R, Monrad RN. Enzymatic degradation of lignin-carbohydrate complexes (LCCs): Model studies using a fungal glucuronoyl esterase from *Cerrena unicolor*. Biotechnol Bioeng 2015;112:914-922.

[27] Katsimpouras C, Bénarouche A, Navarro D, Karpusas M, Dimarogona M, Berrin J-G, *et al.* Enzymatic synthesis of model substrates recognized by glucuronoyl esterases from *Podospora anserina* and *Myceliophthora thermophila*. Appl Microbiol Biotechnol 2014;98:5507-5516.

[28] Vafiadi C, Topakas E, Biely P, Christakopoulos P. Purification, characterization and mass spectrometric sequencing of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*. FEMS Microbiol Lett 2009;296:178-184.

[29] Huynh HH, Arioka M. Functional expression and characterization of a glucuronoyl esterase from the fungus *Neurospora crassa*: identification of novel consensus sequences containing the catalytic triad. J Gen Appl Microbiol 2016;87:1765-1772.

[30] Topakas E, Moukouli M, Dimarogona M, Vafiadi C, Christakopoulos P. Functional expression of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*: identification of the nucleophilic serine. Appl Microbiol Biotechnol 2010;87:1765-1772.

[31] Agger JW, Busk PK, Pilgaard B, Meyer AS, Lange L. A new functional classification of glucuronoyl esterases by Peptide Pattern Recognition. Front Microbiol. 2017;8:309.

[32] Chen Y, Black DS, Reilly PJ. Carboxylic ester hydrolases: Classification and database derived from their primary, secondary, and tertiary structures. Protein Sci 2016;25:1942-1953.

[33] De Santi C, Willassen NP, Williamson A. Biochemical characterization of a family 15 carbohydrate esterase from a bacterial marine arctic metagenome. PLoS One 2016;11:e0159345.

[34] Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 2014;42:W320-W324.

[35] Hüttner S, Klaubauf S, de Vries RP, Olsson L. Characterisation of three fungal glucuronoyl esterases on glucuronic acid ester model compounds. Appl Microbiol Biotechnol 2017; 101:5301-5311.

[36] Biely P, Malovíková A, Uhliariková I, Li X-L, Wong DWS. Glucuronoyl esterases are active on the polymeric substrate methyl esterified glucuronoxylan. FEBS Lett 2015;589:2334-2339.

Figure captions

Fig. 1 Model structure of 4-*O*-methyl-D-glucurono(arabino)xylan [modified from 12, 15]. GE indicates glucuronoyl esterase. In nature *O is typically linked to lignin instead of methyl group.

Fig. 2 Phylogenetic relationships among the (putative) fungal GEs. △, characterized GEs. The sequences used for BLASTP search in genome mining analysis are indicated as ◇ for characterized GE and ○ for putative GEs. Filled symbols indicate selected GEs for characterization in the present study. Bact. indicates group of bacterial GEs. U indicates ungrouped sequences. Feruloyl esterases (FAEs) were used as an outgroup. Full phylogenetic tree is given in A.1 in Supplementary materials. Full enzyme names, details and sequences are given in Table 2 and A.2 in Supplementary materials.

Table 1 Characterized GEs with their properties

Origin	Enzyme	Sub-group	Production[a]	Molecular mass (kDa)[2]	pH		Temperature (°C)		p*I*[b]	Reference
					Optimum	Stability	Optimum	Stability		
Fungi										
Schizophyllum commune	ScGE	4	Pur	44	7.0	-	50	-	3.5	[10]
	(rScGE)		HP	53 (42)	-	-	-	-	3.7	[23]
Hypocrea jecorina (Trichoderma reesei)	Cip2	5	HT	55	5.5	4.0-8.0	40-60	<40	7.9	[12]
Phanerochaete chrysosporium	PcGE1	4	HAv, Pc, Sc	47	5.0-6.0	-	45-55	-	6.5 (5.5)	[25, 35]
Phanerochaete chrysosporium	PcGE2	4	HSc	42	5.0-6.0	-	45-55	-	4.7 (4.8)	[25]
Myceliophthora thermophila (Sporotrichum thermophile)	StGE1	8	Pur	58	6.0	7.0-8.0	60	<55	6.7	[28]
Myceliophthora thermophila (Sporotrichum thermophile)	StGE2	1	HP	43	7.0	4.0-10.0	55	<50	(5.8)	[30]
Phanerochaete carnosa	PcGCE	4	HP, HAt	72 (42)	6.0	-	40	-	-	[24]
Podospora anserina	PaGE1	5	HP	63	-	-	-	-	7.6 and 8.2(6.9)	[27]
Cerrena unicolor	CuGE	4	HAo	58 (48)	-	-	-	-	-	[26]
Neurospora crassa	NcGE	8	HP	44	7.0	4.0-7.0	40-50	<70	-	[29]
Acremonium alcalophilum	AaGE1	5	HP	72 (53)	-	7.0-11.0	-	<50	-	[35]
Wolfiporia cocos	WcGE1	4	HP	45 (44)	-	7.0[c]	-	<40	-	[35]
Bacteria										
Ruminococcus flavefaciens	cesA	-	HE	46	-	-	-	-	-	[36]

uncultured bacterium	MZ0003	-	HE	46	8.0	7.0-9.5	35	<30	-	[33]

[a] Pur, purified from the original source; H, homologous expression (Ao, *Aspergillus oryzae*; At, *Arabidopsis thaliana*, Av, *Aspergillus vadensis*; E, *Escherichia coli*; P, *Pichia pastoris*; Pc, *Pycnoporus cinnabarinus*; T, *Trichoderma reesei*; Sc, *Schizophyllum commune*)

[b] Parentheses indicate calculated values

[c] pH stability varied on the buffer [35]

Table 2 Molecular mass, production level and specific activity of characterized GEs in this work[a].

Fungal species	Accession number	Sub-group	Name[b]	Calculated molecular mass (kDa)	Apparent molecular mass (kDa)	Deglycosylated protein (kDa)	Calculated p*I*	Production (mg/L)	Relative activity[c] (nkat/mg)
Podospora anserina	CAP59671	1	PaGE2	44.3	nd	nd	8.3	np	na
Leptosphaeria maculans	CBX90574	1	LmGE1	41.6	nd	nd	8.2	np	Active[d]
Dichomitus squalens	jgi|Dicsq1|58498	1	DsGE1	41.8	65-70	45	5.0	2	1,159
Coprinopsis cinerea	jgi|Copci1|5044	2	CcGE1	43.4	60-75	45	6.0	52	na
Penicillium rubens	CAP91804	3	PrGE1 (Pc13g07350)	40.2	50-60	42	6.3	14	162
Agaricus bisporus	jgi|Agabi_varbisH97_2|209748	3	AbGE1	46.5	nd	nd	5.9	np	na
Hypholoma sublateritium	jgi|Hypsu1|50423	3	HsGE1	47.4	48.8	nd	5.6	336	2,334
Schizophyllum commune	XP_003026289	4	**ScGE**	40.2	40	36	4.3	25	4
Phanerochaete carnosa	AFM93784	4	**PcGCE**	42.5	nd	nd	4.6	330	4,501
Stereum hirsutum	jgi|Stehi1|96554	4	ShGE1	47.2	nd	nd	4.6	296	333
Dichomitus squalens	jgi|Dicsq1|107426	4	DsGE2	46.7	75-100	50	4.2	38	46
Podospora anserina	XP_001903136	5	**PaGE1**	49.2	60	60	8.1	26	60
Trichoderma reesei	AAP57749	5	**Cip2**	46.7	71.4	nd	6.4	323	333
Ascobolus immersus	jgi|Ascim1|226781	5	AiGE1	47.6	nd	45.5	7.7	<1	225
Piriformospora indica	CCA74892	5	PiGE1	47.9	nd	48.9	8.3	5	67
Botryosphaeria dothidea	jgi|Botdo1|13681	6	BdGE1	39.0	nd	42.4	7.2	1	77
Stagonospora nodorum	jgi|Stano2|2908	7	SnGE1	39.5	nd	nd	7.8	np	Active[d]
Myceliophthora thermophila (Sporotrichum thermophile)	AEO60464.1	8	**StGE1**	40.3	40	40	5.6	66	31

Podospora anserina	CAP65970	8	PaGE3	40.3	50-60	40	8.5	44	36
Apiospora montagnei	jgi|Apimo1|126025	8	AmGE1	40.0	40-50	40	6.7	69	17
Piromyces sp. E2	jgi|PirE2_1|60981	U	PirGE1	40.6	42	40	6.1	<1	520

[a] nd, not detected; np, not produced; na, not active

[b] Name in bold indicates the previously reported GEs

[c] The assay performed at 45°C using 2 mM benzyl-D-glucuronic acid ester in 73 mM phosphate buffer, pH 6.0

[d] The enzyme was active but specific activity cannot be calculated because of undetectable protein level on the SDS-PAGE