

**Unraveling the diversity within
CAZy families related to
hemicellulose degradation**

Xinxin Li
2022.09

Xinxin Li

Unraveling the diversity within CAZy families related to hemicellulose degradation
PhD thesis, Utrecht University, Utrecht, the Netherlands (2022)

ISBN: 978-94-6423-969-0

Cover and invitation design: Xinxin Li

Lay-out design: Xinxin Li

Printing: ProefschriftMaken || www.proefschriftmaken.nl

Copyright ©2022 by Xinxin Li

All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system or transmitted in any other means, without the permission of the author, or when appropriate of the publisher of the represented published articles.

Unraveling the diversity within CAZy families related to hemicellulose degradation

**Het ontrafelen van de diversiteit binnen CAZy families
met betrekking tot de afbraak van hemicellulose**
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 21 september 2022 des middags te 4.15 uur

door

Xinxin Li

geboren op 29 juni 1992
te Henan, China

The research described in this thesis was performed at the Fungal Physiology Group, Westerdijk Fungal Biodiversity Institute (formerly CBS-KNAW), Utrecht, the Netherlands and supported by the China Scholarship Council (grant no. 201803250066).



WESTERDIJK
FUNGALBIO
DIVERSITY
INSTITUTE



Promotor:

Prof. dr. ir. R.P. de Vries

Copromotor:

Dr. M.A. Kabel

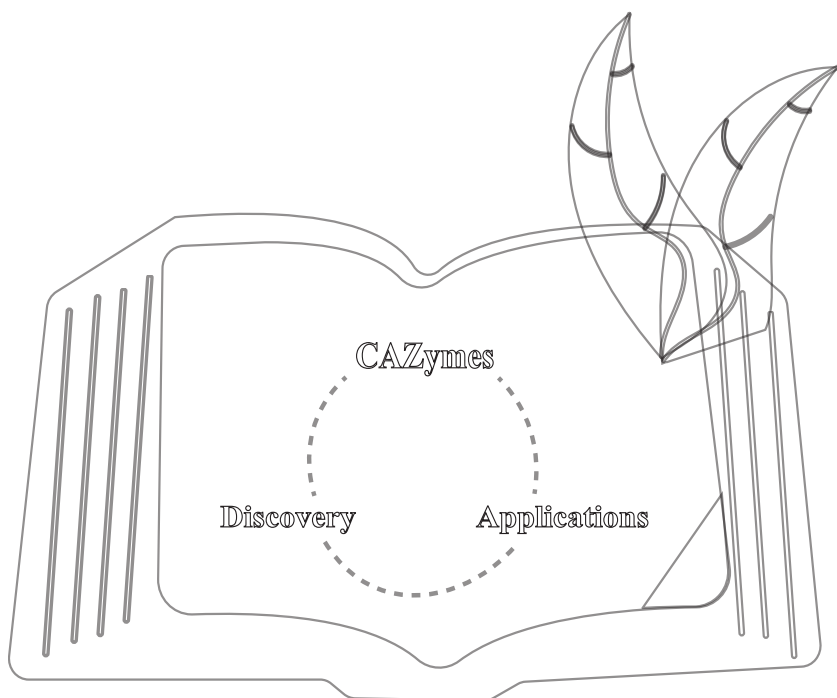
**This book is dedicated to my beloved parents and brother
Thanks for always being there for me and supporting me every step of
the way on this journey**

CONTENTS

Chapter 1	General introduction	8
Chapter 2	Functional validation of two fungal subfamilies in Carbohydrate Esterase family 1 by biochemical characterization of esterases from uncharacterized branches	32
Chapter 3	Glycoside Hydrolase family 30 harbors fungal subfamilies with distinct polysaccharide specificities	52
Chapter 4	Comparative characterization of nine novel GH51, GH54 and GH62 α -L-arabinofuranosidases from <i>Penicillium subrubescens</i>	74
Chapter 5	GH10 and GH11 endoxylanases in <i>Penicillium subrubescens</i> : comparative characterization and synergy with GH51, GH54, GH62 α -L-arabinofuranosidases from the same fungus	88
Chapter 6	Fungal Glycoside Hydrolase family 44 xyloglucanases are restricted to the phylum Basidiomycota and show a distinct xyloglucan cleavage pattern	112
Chapter 7	Summary and general discussion	136
Appendix	English summary	156
	Nederlandse samenvatting	158
	Curriculum Vitae	160
	List of publications	161
	Acknowledgements	162

章节一

综述



Chapter 1

General introduction

Part of this chapter was published in Bioresource Technology

Li, Xinxin, Adiphol Dilokpimol, Mirjam A. Kabel, and Ronald P. de Vries

Volume 344, January 2022, 126290

<https://doi.org/10.1016/j.biortech.2021.126290>

Introduction

Lignocellulosic biomass, including agricultural waste streams, forest residues, and short rotation coppices, is the most abundant resource on earth. This lignocellulose has become an important material of the production of alternative fuels (e.g., bioethanol, biomethanol) for most countries in the world, especially in developed countries [1-3]. In addition to the production of biofuels, lignocellulosic biomass is also a feedstock for animal feed, textiles and pulp & paper [4]. Polysaccharides (cellulose, hemicellulose, pectin) and the aromatic polymer lignin are the main components of the cell walls of lignocellulosic biomass [5] (**Figure 1**). Cellulose is a long linear polymer composed of D-glucosyl residues, while hemicellulose is a heteropolysaccharide made up of hexoses (D-glucosyl, D-mannosyl, and D-galactosyl residues) and pentoses (D-xylosyl and L-arabinosyl residues) [6, 7]. Xylan, xyloglucan and galactomannan are the main hemicellulose structures, with xylan being the most common abundant hemicellulose in a wide range of plants, e.g., wood, monocots, and dicots [8, 9].

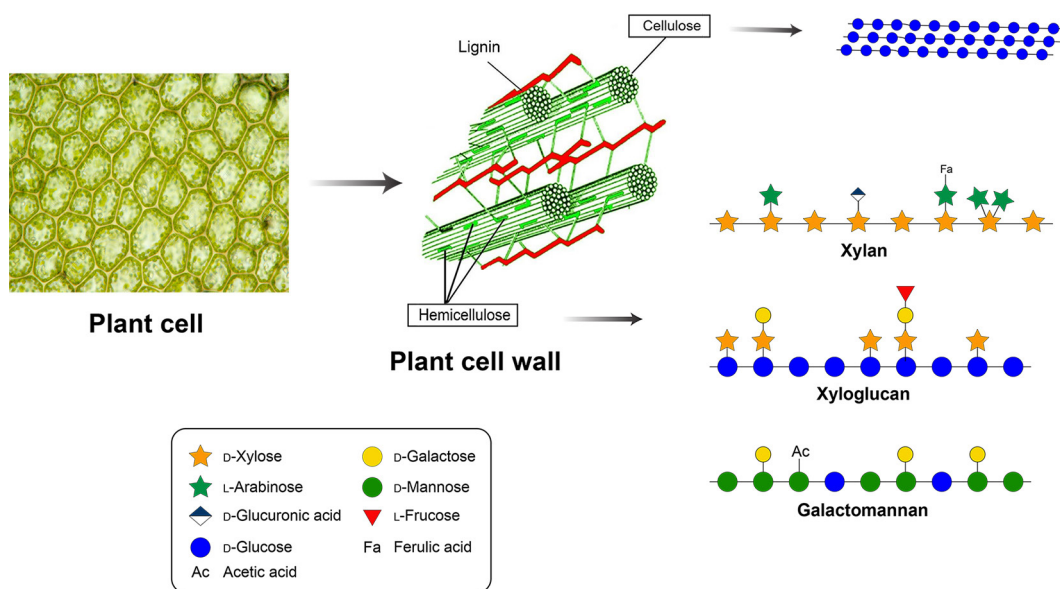


Figure 1 Structure of plant cell wall and schematic representation of polysaccharides. Polysaccharide structures were drawn based on [14] and [15].

Xylan forms a tightly bound structure around the crystalline cellulose together with lignin, which ensures the structural integrity of plants and provides resistance to pathogenic attack, pests and to enzymatic degradation. This complex and recalcitrant architecture of lignocellulose is the key factor that limits its large-scale use [10]. Targeting xylan degradation can to some extent overcome the recalcitrance of lignocellulose and is therefore significant for widespread utilization of lignocellulosic biomass in industry. Since the end of the 1980's, enzymatic processes have been widely applied in the industrial processing of lignocellulose to replace the use of various toxic chemicals, as enzymes are more eco-friendly and have high specificity and efficiency [11-13]. The popularity of enzymes in industry has led to an increasing demand for enzyme cocktails in industry.

To improve the use of lignocellulose by efficient enzymatic degradation of xylan, in this general introduction, we provide an overview of the major enzymes involved in xylan degradation, and their substrate specificities. In addition, methodological advances in the discovery of new enzyme candidates for xylan degradation are discussed, with emphasis on the strategy of fungal genome mining. These discovery strategies of xylan-degrading enzymes are also applicable to discover other

enzymes, such as those involved in the degradation of xyloglucan, arabinogalactan. Finally, we review the current applications of enzymes in industry.

Categories of xylan and their structures

Xylan generally consists of a β -(1 \rightarrow 4)-linked D-xylosyl backbone, which is further decorated by different residues, i.e., α -L-arabinose, α -D-glucuronic acid, 4-*O*-methyl-glucuronic acid (MeGlcA), acetic acid [9]. In addition, linear β -(1 \rightarrow 4)-homoxylan without decorations has been isolated successively from esparto grass [16], guar seed husk [17], and tobacco stalks [18]. Furthermore, xylan with mixed β -(1 \rightarrow 3)(1 \rightarrow 4)-linked D-xylosyl units as backbone has also been found in *Plantago* species [19]. Since β -(1 \rightarrow 4)-homoxylan and β -(1 \rightarrow 3)(1 \rightarrow 4)-linked xylan are rarely found in plants, we mainly focus on β -(1 \rightarrow 4) linked xylan with different decorations. Although xylan content, type and degree of xylan substituents vary amongst different botanical species and tissue types [20], generally two main types are recognized: glucuronoxylan (GX, *O*-acetyl-4-*O*-methylglucuronoxylan), and glucuronoarabinoxylan (GAX, (*O*-acetyl)-arabino-4-*O*-methylglucuronoxylan) [20, 21] (**Figure 2**).

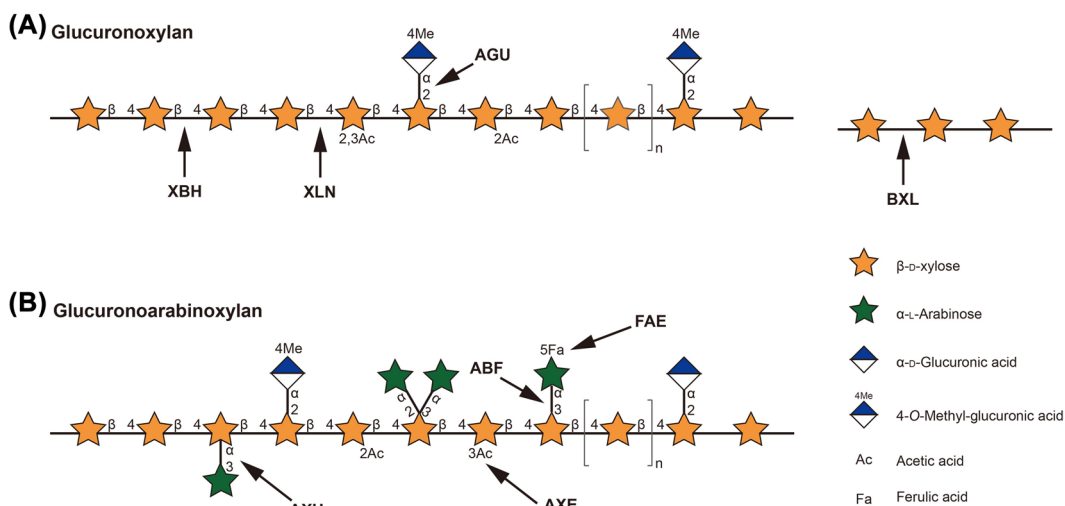


Figure 2 Schematic representation of xylans and corresponding xylanolytic enzymes. (A), glucuronoxylan; (B), glucuronoarabinoxylan. Enzyme abbreviations are presented in Table 1. Polysaccharide structures were drawn based on [21] and [24].

GX is mainly found in hardwoods, accounting for 10-35% of the total dry mass [20, 21]. Around 10% of the D-xylosyl residues are *O*-2 substituted with MeGlcA residues, and about 35-70% are *O*-2 and/or *O*-3 linked with acetic acid [20]. Acetylation has been found to occur more frequent at the *O*-3 than at the *O*-2 positions of the D-xylosyl residues [22]. GAX is the representative xylan in softwoods (5-10% of dry weight) and grasses (or cereals) (20-35% of dry mass) [20, 21, 23]. GAX of softwoods has generally no acetylation [24], around 20% of the D-xylosyl residues are *O*-2 substituted with (Me)GlcA residues, and about 13% are *O*-3 substituted with L-arabinoxyl residues [25]. Whereas, GAX of grasses and cereals are lightly acetylated [26], the amount of L-arabinoxyl residues is relatively high, and for example for corn pericarp GAX the L-arabinose can be further substituted via *O*-2 to one or more carbohydrate residues [25]. GAX in cereals usually contain less or no (Me)GlcA residues, and mainly found in cereal grain endosperm [20]. Up to a ratio of L-arabinoxyl substitutions to D-xylosyl of 0.48 to 1.07 has been found for endospermic xylans from different cereal species [27]. L-Arabinoxyl units in GAX can be further esterified at their *O*-5

position by ferulic acid and, although to a lower extent, by *p*-coumaric acid [20, 21].

Xylan-degrading enzymes and their substrate specificities

Enzymatic hydrolysis of xylan involves a multi-enzyme system (**Table 1**), named xylanolytic enzymes, including main- and side- chain degrading enzymes [28]. Xylanolytic enzymes can be produced by various microorganisms, e.g., fungi, bacteria, and actinomycetes, of which fungi, especially filamentous fungi, are most popular from an industrial point of view [29, 30]. These enzymes from fungi are nearly all included in the CAZy database, in which they are classified into different classes and further divided into different families based on their amino acid sequences, structures and molecular mechanisms [31]. Moreover, some families have been subdivided into different subfamilies (SFs).

Table 1 Overview of fungal xylanolytic enzymes for xylan degradation.

CAZy	Activity	EC number	Abbreviation	Main CAZy family ^b
GH	Endoxylanase	3.2.1.8	XLN	GH10, GH11
	Xylosidase	3.2.1.37	BXL	GH3, GH43
	Xylobiohydrolase	3.2.1.-	XBH	GH30_7
	Arabinofuranosidase	3.2.1.55	ABF	GH43, GH51, GH54, GH62
	Arabinoxylan arabinofuranohydrolase	3.2.1.55	AXH	GH62
	Glucuronidase	3.2.1.131	AGU	GH67, GH115
CE	Acetyl xylan esterase	3.1.1.72	AXE	CE1
	Feruloyl esterase	3.1.1.73	FAE	CE1
	Glucuronoyl esterase	3.1.1.B11	GE	CE15
AA	Lytic polysaccharide monooxygenase	N.A. ^a	LPMO	AA9, AA14 ^c

a, Not categorized by the International Union of Biochemistry and Molecular Biology (IUBMB).

b, Only main families of characterized enzymes from fungi based on the current CAZy annotation are shown.

c, To date, only two AA9 LPMO has been shown to cleave isolated xylan oxidatively (e.g., *LsAA9A* [32] and *PMO9A_MALCI* [33]). For various AA9 and AA14 LPMOs, oxidative xylan cleavage has been observed, but only when xylan was bound to cellulose, resulting in a simultaneous oxidative cellulose cleavage (reviewed in [34]).

Xylan main chain degrading enzymes

β -D-Xylanase (XLN) and β -D-xylobiohydrolase (XBH) are glycoside hydrolases (GHs) able to cleave the xylan backbone, likewise, β -D-xylosidase (BXL) degrade the β -(1 \rightarrow 4)-D-xylosyl backbone, but preferably of oligosaccharides. XLN is typical endo-active, hence, releasing mainly xylooligosaccharides, while XBH and BXL have been recognized as exo-active, releasing solely xylobiose and D-xylose, respectively [35-37]. Fungal XLNs are mainly classified in GH10 and GH11, XBHs in GH30_SF7, and BXLs in GH3 and GH43. XLNs and BXLs have been described in detail in several reviews [38-42], whereas XBH as a novel activity has only recently been discovered and reported [37, 43-46].

GH10 XLNs generally have a broader substrate specificity than GH11 XLNs, as GH10 XLNs tolerate the presence of decorations on the xylan backbone more than GH11 XLNs do [47]. As a result, GH10 XLNs can cleave the glycosidic linkage next to a D-xylosyl residue decorated with one or two residues (e.g., at the non-reducing side), whereas GH11 XLNs only cleave the xylan backbone that has at least three consecutive undecorated D-xylosyl residues, and cannot cleave at the glycosidic linkage next to a branch [38, 48]. BXL is characterized generally using an artificial substrate, i.e., *p*-nitrophenyl- β -D-xylopyranoside (*p*NPXyl). So far, almost all reported GH3 and GH43 BXLs are active against this substrate. Some BXLs were reported to exhibit multiple activities, shown by the

hydrolysis of different substrates in addition to *p*NPXyl, e.g., *p*-nitrophenyl- α -L-arabinofuranoside (*p*NP α Ara) [49-52] and *p*-nitrophenyl- β -D-glucopyranoside (*p*NPGLc) [50]. GH43 BXLs generally do not have the ability to transglycosylate at high substrate concentrations [40, 53], whereas transglycosylation activity has been reported for several GH3 BXLs [50, 54, 55]. This suggests that candidates for typical BXL hydrolase-activity can better be selected from GH43 [40].

XBH, as novel enzyme, has only attracted the attention of researchers in recent years. The earliest XBH activity was suggested for *Penicillium purpurogenum* 'arabinofuranosidase ABF3' of GH43, since it released more xylobiose than D-xylose when X3 up to X5 were used as substrates. However ABF3 released different xylooligosaccharides when incubated with arabinoxylan (AX) [43]. However, the subsequently reported XBHs were all from SF7 of GH30, and almost all of them showed a preference for less decorated GX, compared to highly substituted GAX [37, 44-46]. As an exception, *Acremonium alcalophilum* XBH (*AaXyn30A*) exhibited the highest activity towards rhodymenan, which is a linear β -(1 \rightarrow 3)(1 \rightarrow 4)-xylan, followed by GX [37]. To date, only a limited number of fungal XBHs have been characterized, which to some extent hinders to understand the diversity of XBH in terms of substrate specificity.

Xylan side chain degrading enzymes

Xylan side chain degrading enzymes are able to remove side branches and decoration groups in the different types of xylan, mainly including α -L-arabinofuranosidase (ABF), α -glucuronidase (AGU), acetyl xylan esterase (AXE), feruloyl esterase (FAE), and glucuronoyl esterase (GE). They belong to different classes, i.e., GH and carbohydrate esterase (CE).

ABF is an exo-acting enzyme of the GH class, which hydrolyzes the linkages between L-arabinosyl residues and D-xylosyl units of the xylan backbone to release L-arabinose. Some ABFs also degrade pectin side chains and other L-arabinosyl-containing oligosaccharides [49, 56, 57]. In the CAZY database, the fungal ABFs mainly belong to GH43, GH51, GH54 and GH62 [31]. According to their mode of action towards substrates, ABFs are divided into two groups. ABFs in group A are able to remove L-arabinosyl residues decorated at *O*-2 or *O*-3 position of mono-substituted D-xylosyl residues, e.g., *Myceliophthora thermophila* C1 Abf3 in GH51 [58]. In contrast, ABFs in group B can selectively release *O*-3 L-arabinose from doubly arabinosylated D-xylosyl residues (*O*-2 and *O*-3), e.g., *M. thermophila* C1 Abn7 in GH43 [58]. In addition, based on substrate preference, ABFs have been classified into four types (1, 2, 3, 4) [59]: type 1, not active towards polymers; type 2, active towards polymers; type 3, specific for arabinoxylan; type 4, not active on the synthetic substrate *p*NP α Ara. Previously, type 1 and 2 were assigned to GH51 and GH54 ABFs, respectively, and type 3 corresponded to GH43 and GH62 ABFs [60]. However, as additional enzymes were characterized, the drawbacks of this classification based on substrate specificity became evident, because multiple types of ABF were present in the same family. For example, in GH51 corresponding to type 1, some ABFs, like *Aureobasidium melanogenum* ATCC 20524 AbfB [61] and *P. chrysogenum* 31B AFQ1 [56], were also active on AX. ABFs have proven to degrade different substrates, e.g., *p*NP α Ara, sugar beet arabinan and/or AX [49, 57, 62-64]. Enzymes that specifically hydrolyze linkages between L-arabinosyl and D-xylosyl residues in AX are also named arabinoxylan arabinofuranohydrolases (AXHs) and are mainly from GH62 [64-66].

AGU is another key xylan side chain degrading enzyme in the GH class, which catalyzes the breakdown of α -(1 \rightarrow 2)-glycosidic linkages between (Me)GlcA and D-xylosyl residues in glucurono-xylooligosaccharides or GX. Fungal AGUs are classified into GH67 and GH115. Fungal GH115 AGUs release MeGlcA by the hydrolysis of α -(1 \rightarrow 2)-glycosidic bond between MeGlcA and mono-substituted internal- and end D-xylosyl units, of polymeric GX and/or xylan oligomers [67-70]. However, fungal AGUs from GH67 reported cannot operate on polymeric substrates, and exclusively hydrolyze the α -(1 \rightarrow 2)-glycosidic bond between MeGlcA and terminal, non-reducing end, D-xylosyl residues of xylooligosaccharides [69, 71-73]. As an exception, *A. nidulans* AguA was also active on larch wood xylan [49]. So far, only a limited number of fungal AGU has been characterized.

AXE belongs to the CE class, which catalyzes the hydrolysis of ester linkages between the acetyl groups and the xylan backbone to release acetic acid. Common model substrates used for detection of AXE activity are including *p*NP-acetate, 4-methylumbelliferyl acetate (MUB-acetate), α -naphthyl acetate, and 2,3,4,5-tetra-*O*-acetyl-D-xylose, of which the first former three are commercially available. Apart from that, AXE is also active on natural substrates, e.g., acetylated xylan and xylooligosaccharides. The characterized AXEs from fungi are classified into CE1, CE2, CE3, CE4, CE5, CE6 and CE16. CE1 is the largest family which contains most characterized fungal AXEs, while CE4 and CE5 mainly contained chitin deacetylases and cutinases, respectively. Currently, there are only few characterized fungal enzymes from CE2, CE3 and CE16. The specificity of AXEs from different families varies significantly towards natural substrates. CE1 AXEs regioselectively cleave the substituents in the *O*-2 and *O*-3 position, and deacetylate the *O*-2 position faster than the *O*-3 position [74]. CE4 AXEs prefer to deacetylate oligosaccharides with longer chain, which cannot attack di-acetylated D-xylosyl residues [28, 75]. Altaner, Saake [74] reported that CE4 AXEs seem to interact only with acetyl groups at the *O*-3 position of D-xylosyl units, whereas CE5 AXEs attack solely the *O*-2 position. CE6 AXEs have a broad specificity, which are able to target *O*-2 mono-acetylated D-xylosyl residues as well as *O*-2 and *O*-3 di-acetylated D-xylose [76]. AXEs from CE16 showed various specificity of deacetylation. For example *Myceliophthora thermophila* MtAcE of CE16 preferred to release the acetyl group at the *O*-2 position, similar to CE1 AXEs [77], whereas *Trichoderma reesei* TrCE16 had a preference for removing the acetyl group of D-xylosyl units at the *O*-3 and *O*-4 positions in oligosaccharides of acetyl-glucuronoxylan [78]. *A. niger* AnCE16A tended to rapidly attack *O*-2 and *O*-3 di-acetylated D-xylose in polymeric xylans, followed by *O*-2 and *O*-3 mono-acetylated D-xylose, whereas, it poorly hydrolyzed acetyl groups of the *O*-2 D-xylosyl position in methyl β -D-xylopyranoside diacetates and triacetates [79]. The mode of action of CE2 and CE3 AXEs still needs to be studied.

FAE hydrolyzes the ester bonds between hydroxycinnamic acids (e.g., mostly (di-) ferulic acid and, to a minor extent, coumaric acid) and L-arabinosyl residues in xylan. Methyl ferulate (MFA), methyl sinapate (MSA), methyl caffeate (MCA), and methyl *p*-coumarate (MpCA) are most often used as model substrates to determine FAE activity. Based on activity towards these model substrates, FAEs were classified into four types (A, B, C and D) [80]. Type A and B are inactive towards MCA containing methyl 3,4-dihydroxycinnamate and MSA containing methyl 3,5-dimethoxy-4-hydroxycinnamate, respectively, while type C and D can hydrolyze all four model substrates. Unlike type C FAEs, type D FAEs are also capable of releasing 5-5'-diferulic acid from natural substrate. However, the characterization of additional FAEs revealed that this classification system cannot classify all the known FAE consistently. Recently, Dilokpimol, Mäkelä [81] introduced a refined FAE classification based on phylogenetic analysis of available fungal genome. This refined classification separated FAEs into 13 subfamilies, which not only classifies different types of FAE, but also reflects the evolutionary relationship among different FAEs [81]. Only two SFs (SF5 and SF6) are classified as CE1 in the CAZy database, which are related to AXE. Other SFs are not in the CAZy database [31]. These FAEs are related to tannases (SF1-4 and SF9-11), lipases (SF7), and lipases and choline esterases (SF12 and SF13). This demonstrates that FAEs are evolved from highly diverse esterase families [81, 82].

GE belongs to CE15, which catalyzes the hydrolysis of ester bonds between aliphatic alcohols in lignin and the MeGlcA side chains of glucuronoxylan [83]. Benzyl D-glucuronate is the commercially available substrate for screening of GE activity, while others, e.g., benzyl methyl α -D-glucopyranosiduronate, benzyl methyl 4-*O*-methyl- α -D-glucopyranosiduronate and lignin-carbohydrate complexes (LCCs) are prepared by specific methods. GEs not only synergistically act with xylanases, but also potentially with other lignocellulolytic enzymes [84, 85].

Lytic polysaccharide monoxygenases

In recent years, it has been reported that lytic polysaccharide monoxygenase (LPMO) in the class of auxiliary activity (AA) also contribute to xylan degradation (i.e., from AA9 and AA14)

[32, 33, 86-88]. AA9 is the largest (fungal) LPMO family, and these LPMOs have been shown to oxidatively degrade cellulose. A selection of AA9 LPMOs has been shown to also oxidatively cleave xyloglucan, while so far only limited LPMOs are described to be active towards xylan [89]. *Lentinus similis* LsAA9A [32] and *Malbranchea cinnamomea* rPMO9A_MALCI [33] were reported to oxidatively cleave isolated xylan substrates. Other xylan-active LPMOs in AA9, such as *Malbranchea cinnamomea* McAA9s (McAA9A, McAA9B, McAA9F, McAA9H) [86] and *Myceliophthora thermophila* C1 MtLPMO9A [87], oxidatively cleaved only cellulose-associated xylan. For AA14 LPMOs, Couturier, Ladeveze [88] characterized the first two xylan active LPMOs, i.e., PCAA14A and PCAA14B from *Pycnoporus coccineus*, which acted on xylan coated crystalline cellulose, whereas they were inactive towards pure xylan. Remarkably, the combination of LPMOs and GHs showed a strong synergistic effect on xylan degradation [90, 91]. Similar synergistic effects were also found between xylan main- and side chain degrading enzymes, and even between different side chain degrading enzymes [75, 91, 92]. Therefore, complete hydrolysis of xylan is achievable only through a rational combination of multiple enzyme activities.

Different strategies for the discovery of enzymes

Fungi are of interest because they are able to secrete a complete enzyme system required for their growth habitat [30, 93, 94]. Recently, the global fungal diversity is estimated to range between 2.2 and 3.8 million species [95], indicating a rich resource of enzymes to be exploited. However, this requires efficient and sensitive screening strategies. Methodological advances in the discovery of new enzyme candidates from fungi are discussed in this section.

Strain screening

Strain screening is a traditional and powerful approach for discovery of new enzymes [96]. It is based on microbial diversity and versatility, using a well-considered screening system to examine as many microorganisms as possible, followed by culturing and subsequent screening of pure organisms to obtain the desired enzymes [96]. Currently, many research institutes and companies have established collections of microorganisms from a variety of environments, including “common environments” (e.g., soil or seawater) and “extreme environments” (e.g., hot springs, Antarctic ice, alkaline lakes, etc.) [97]. Some microbial arrays have been cultured and screened, from which excellent candidates have been selected based on their ability to provide enzymes with new properties [98-101]. For example, Pedersen, Lauritzen, Frisvad, et al. [98] evaluated the production and activity of enzymes from 20 *Aspergillus* strains, of which enzymes with high XLN activity were found from *A. brasiliensis*, *A. niger* and *A. japonicus*. In addition, in *A. brasiliensis* and *A. niger* strains, thermostable BXL with high temperature optimum (~ 75°C) were also found [98]. Similarly, Sridevi, Charya [102] cultivated 30 different fungi from various genera (e.g., *Alternaria*, *Aspergillus*, *Curvularia*, *Drechslera*, *Fusarium*, *Penicillium*, *Rhizopus*, *Trichoderma*). From the genus *Aspergillus*, some cellulase-free XLN with high production levels were reported [102]. In addition, enzymes with similarly promising industrial applications were also found in *Fusarium* CDC-8, combining high XLN activity with low cellulase production [102].

Although strain screening is a useful method to find new enzymes, more than 99% of microorganism in most environments are not suitable for culture, and little is known about their genomes, genes and encoded enzyme activities [103]. The organisms that can be cultured encompass only a tiny fraction of the existing genomic diversity in any given environment. Screening only these organisms would limit the potential spectrum of discovering new enzymes for the bioprocessing industry. In addition, fungal genome sequencing demonstrated that a large number of the genes from the genomes of culturable microorganisms have no assigned function, because of their low (or no) homology to enzymes and/or proteins that have been characterized [103, 104]. Thus, mining the resources of non-cultured microbes, and studying the diversity of proteins that have not yet been characterized in cultured microbes can also contribute to the discovery of new enzymes.

Table 2 Comparison of the number of genes from selected fungal in the related xylan degrading enzyme families in the CAZy database. The data was based on the CAZy annotation from published genomes (until September 2021) in MycoCosm [117]. Values in bold represent the highest number of genes per fungal strain found per CAZy family.

Phylum	Xylanolytic enzymes		XLN		XBH		BXL		ABF			AXH		AGU		AXE / FAE		GE		LPMO		
	Families		GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH
	Species ^a		10	11	30	7	3	43	51	54	62	67	115	1	1	15	9	14	AA	AA	AA	AA
			2	3	0	0	19	11	4	1	1	1	0	3	0	7	0					
	<i>Aspergillus niger</i>		4	4	0	0	23	20	3	1	2	1	4	3	0	8	0					
	<i>Aspergillus oryzae</i>		3	2	0	0	21	19	2	1	2	1	1	3	0	10	0					
	<i>Aspergillus nidulans</i>		2	3	0	0	20	13	4	1	1	0	0	1	0	5	1					
	<i>Aspergillus japonicus</i>		4	3	0	0	19	18	2	1	2	1	1	2	1	7	1					
	<i>Aspergillus fumigatus</i>		2	4	0	0	16	11	4	1	1	0	0	1	0	6	0					
	<i>Aspergillus luchuensis</i>		1	3	2	2	13	2	0	2	1	1	1	0	1	3	1					
	<i>Trichoderma reesei</i>		3	1	0	0	18	14	3	1	1	0	0	1	1	4	0					
	<i>Penicillium chrysogenum</i>		3	7	0	0	25	30	6	4	4	4	3	3	0	8	1					
Ascomycota	<i>Penicillium subrubescens</i>		8	6	2	11	12	1	0	2	1	3	13	3	32	1						
	<i>Podospora anserina</i>		4	2	0	0	11	7	1	1	0	1	1	6	1	14	0					
	<i>Neurospora erassa</i>		1	0	0	0	2	2	1	0	1	0	1	0	0	4	0					
	<i>Cochliobolus lunatus</i>		2	1	0	0	16	1	3	1	1	0	1	3	0	1	0					
	<i>Mycosphaerella graminicola</i>		2	1	0	0	16	11	3	1	1	0	1	1	0	2	0					
	<i>Zymoseptoria pseudotritici</i>		2	2	0	0	16	6	3	1	1	0	1	1	0	10	0					
	<i>Botrytis cinerea</i>		4	8	2	13	13	2	0	2	1	1	6	2	23	0						
	<i>Myceliophthora thermophila</i>		2	0	0	0	6	1	0	0	0	0	0	0	0	6	1					
	<i>Tuber borchii</i>																					

a. Species with good quality genome sequences and taxonomic diversity are presented in this table.

Table 2 Comparison of the number of genes from selected fungal in the related xylan degrading enzyme families in the CAZy database. The data was based on the CAZy annotation from published genomes (until September 2021) in MycoCosm [117]. Values in bold represent the highest number of genes per fungal strain found per CAZy family (continued).

Phylum	Xylanolytic enzymes		XLN		XBH		BXL		ABF			AGU		AXE/FAE		GE		LPMO		
	Families	Species ^a	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH	GH
			10	11	30	7	3	43	51	54	62	67	115	1	15	9	14	AA	AA	
			5	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	1	0
			2	1	1	3	4	4	2	0	1	0	1	1	0	0	0	0	0	0
			2	2	1	8	4	4	1	0	0	0	2	1	1	11	2	1	11	2
			6	6	1	7	4	4	1	0	3	0	1	4	17	34	5	17	34	5
			5	0	0	8	7	7	2	0	0	0	2	1	2	16	4	2	16	4
			0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
			6	1	1	10	4	4	2	0	0	0	1	4	2	16	2	2	16	2
			3	2	2	13	8	8	3	0	1	0	1	2	1	29	2	1	29	2
			3	0	0	6	1	1	1	0	0	0	1	0	1	2	2	1	2	2
			6	0	0	8	3	3	2	0	0	0	2	3	2	17	4	2	17	4
			6	0	1	13	3	3	2	0	0	0	2	3	2	18	4	2	18	4

a, Species with good quality genome sequences and taxonomic diversity are presented in this table.

Metagenomics

Metagenomics is an effective approach to obtain new enzymes by mining microbial diversity [105]. This approach does not require culturing of microorganisms, as DNA is directly isolated from environmental samples and cloned into suitable vectors to construct so-called “metagenomic” libraries [106, 107]. These metagenomic libraries can be used to identify new genes based on either function (activity) or sequence homology [108, 109]. Function-based screening is a straightforward approach to obtain genes with desired functions by screening and evaluating proteins produced in expression hosts. Sequence-based screening is the use of either hybridization or PCR procedures to obtain genes having homologies to the probe sequence. Both screening methods have accelerated the discovery process of new useful genes from microorganisms. As a result, various new enzymes with unique activities and/or sequences have been discovered [105, 108-116]. For example, in a soil-derived metagenomic library, Hu, Zhang, Li, et al. [116] identified a new gene encoding a GH10 XLN (XynH), and produced the protein using *Escherichia coli* as heterologous expression host. The functional characterization showed that XynH had a lower temperature optimum and a weakly alkaline pH optimum, which differs from most GH10 XLNs. In the metagenomes of natural microbial biomass decay communities, Li, Taghavi, McCorkle, et al. [115] characterized four new GHases. Two of them were highly active on pNP α Ara, while the other two were highly active against pNPGlc and pNPXyl, respectively. Similarly, in a compost microbial metagenomic library, Matsuzawa, Kaneko, Yaoi [114] isolated a GH43 BXL/ABF(CoXyl43), which could enhance the saccharification of different xylan containing plant biomasses, such as rice straw and *Erianthus*.

Compared to strain screening, the metagenomic strategy can avoid microbial isolation and culture processes, allowing exhaustive screening of microbial genomes in their natural environments.

Genome mining

Genome mining is another approach to discover new enzymes based on microbial diversity, which has received increasing attention in recent years. Especially, since the improvements in sequencing technologies resulted in an increasing number of fungal genomes that have been published. These available genomes uncovered a large number of genes encoding CAZymes, which have greatly enhanced our understanding of the diversity of fungi in terms of plant cell wall degradation.

As an illustration, **Table 2** shows a comparison of CAZymes involved in plant polysaccharides degradation of 28 fungal genomes from MycoCosm [117], focusing on the gene number encoding xylanolytic enzymes. This table shows that fungi from different phyla, i.e., Ascomycota and Basidiomycota, differ significantly in the number of genes encoding xylanolytic enzymes. On average, Ascomycota has higher gene copies in GH3, GH11, GH43, GH54, GH62 and GH67 enzymes than Basidiomycota, while the number of AA14 genes in Ascomycota is less than in Basidiomycota. In addition, some genes encoding specific family present phyla specificity. For example, GH54 ABF and GH67 AGU encoded genes are only present in Ascomycota, which might indicate that these enzymes play an essential role in Ascomycota [70]. However, gene numbers are even variable within the same fungal genus. A comparison of gene numbers in *Penicillium* shows that *P. subrubescens* contains significantly more genes than *P. chrysogenum* in most of the CAZy families related to xylan degradation [118]. A similar phenomenon also occurs in some CAZy families (e.g., GH10, GH11 and GH115) of *Aspergillus* [119].

Genes encoding GH10 and GH11 enzymes, which mainly show XLN activity, are present in variable copies from 0 to 8 in the genomes of these selected fungi. The genomes of *Podospira anserina*, *M. thermophila* and *Coprinosia cinerea* contain more copies of XLN than the other fungi, with especially high gene numbers of GH10 XLN in *P. anserina* (eight genes) and GH11 XLN in *M. thermophila* (eight genes). This shows strong expansion of specific families in fungi, probably pointing to diversification within the family in terms of substrate specificity and/or biochemical properties. In addition, genes encoding GH10 XLN are widely present in almost all selected fungi, while genes encoding GH11 XLN are absent in some fungi, e.g., *C. lunatus*, *N. crassa*.

Most of fungi harbor high copies of GH3 and GH43 coding genes, which might be due to the fact that GH3 and GH43 contain various catalytic activities in addition to BXL. e.g., BGL in GH3, ABF / ABN / 1,3GAL (β -(1 \rightarrow 3) galactanase) in GH43 [31]. In addition, species like *P. subrubescens* that have more copies of genes coding xylan side chain degrading enzymes, such as ABF, AXH and AGU than other fungi, showing strong expansions of CAZymes in specific CAZy families (e.g., GH43, GH51) [104]. This feature confirms that *P. subrubescens* is a promising new fungal cell factory for enzyme production, as has been reported previously [120, 121].

For the LPMOs that oxidatively degrade plant biomass, the selected fungal genomes contain highly variable gene numbers (0-34), most of which are from AA9 rather than AA14. AA9 harbors a considerable amount of lytic cellulose monooxygenase. As an exception, the *L. bicolor* genome contains four putative AA14-encoding genes, but no AA9-encoding gene, while no LPMO-encoding genes are present in the genome of *Ustilago maydis*. Furthermore, Basidiomycota genomes generally contain more AA14 genes than Ascomycota genomes, which may be related to the habitat of Basidiomycota. Basidiomycota is reported to live in higher organic-matter forest habitats, which are rich in xylan as the main hemicellulose.

Considering that most genes encoding CAZymes found in fungal genomes are not functionally characterized [104, 122]. Cloning of these unknown genes into a suitable vector, heterologous expression of the recombinant vector in a suitable surrogate host (usually *Escherichia coli* and *Pichia pastoris*), and characterization of the produced proteins using appropriate substrates will likely contribute to the discovery of new activities. Some studies have confirmed the feasibility of the genome mining approach to discover new enzymes. For example, Dilokpimol, Mäkelä [123] confirmed the ability of the genome mining approach to identify fungal GE encoding genes, by successfully obtaining 18 GEs from 21 selected candidates, which were active against benzyl D-glucuronate. Similarly, this author also successfully identified 20 out of 27 putative FAE candidates based on this approach, by testing their activity towards *p*NP-ferulate and/or methyl hydroxycinnamate substrates [124].

Protein engineering

Protein engineering is based on the crystal structure and mechanism of proteins to obtain variants with desired properties by changing potential amino acids. It involves three methods, i.e., rational design, directed evolution and semi-rational design [125, 126]. “Site-directed mutagenesis” is the most classical and effective approach of rational design, which improves enzyme catalytic properties by creating specific mutations in double stranded DNA. For example, the thermostability of *A. niger* BCC14405 XLN at 50°C was greatly improved (18-20-fold) by replacing several surface Ser/Thr residues with Arg in wild type enzyme [127]. Similarly, some XLN mutants from *T. reesei* were obtained by this way, which showed better stability and higher residual activity than wild type enzymes at alkali condition [128]. *De novo* enzyme design is an emerging rational design approach, which depends on different computational algorithms [129, 130]. New enzymes can be either designed by recreating known enzymatic functions in proteins with a different fold or by introduction of activities that have not been observed in natural enzymes before into a chosen protein scaffold [131, 132]. Fenel, Leisola, Jänis, et al. [133] successfully engineered a disulphide bridge into the N-terminal region of *T. reesei* xylanase II (XYNII) by replacing Thr2 and Thr28 with Cys, which increased enzyme thermostability (by about 15°C), as well as the half-life in thermal inactivation of enzyme at 65°C (from 40 s to 20 min) and 70°C (from less than 10 s to 6 min). Rational design is an effective approach but requires an in-depth knowledge of the structure/function relationship in enzymes, which limits the general application of rational design.

Unlike rational design, directed evolution is a robust method based on the Darwinian theory of evolution and does not require an in-depth understanding of structure/function relationship. It has advanced into a standard industrial “tool” to tailor naturally occurring enzymes for various biotechnological applications [134-136]. A typical directed evolution experiment encompasses iterative rounds of gene mutagenesis (e.g., error-prone PCR, DNA shuffling) to construct mutant

libraries, from which desired mutants are obtained through high-throughput screening [135]. For example, to improve the thermostability of xylanase (afxynG1) from *A. fumigatus*, bin Abdul Wahab, bin Jonet, Illias [137] used error-prone PCR to create 5000 afxynG1 mutants and thermally screened them at 60°C for 30 min. Four mutants (T16A/T39I/L176Q, S68R, A60D and Q47P/S159R) with higher catalytic activity than that of wild type were obtained, of which the mutant with three-point mutations also had increased stability. It retained around 50% of its activity at 70°C for 1 h, while the wild type was fully inactive after 50 min incubation [137]. Although directed evolution has shown to be a good approach to obtain new enzymes, the libraries created by this random method are usually huge, which necessitates tedious and time-consuming screening. Creating a good mutant library and developing a suitable screening platform are the most critical parts of all directed evolution exercises, while they have proven to be challenging.

More recently, semi-rational design is an increasingly used method in protein engineering, which combines protein sequence and structural information with advanced computational and machine learning algorithms to design smaller and more focused libraries of protein variants to efficiently obtain the desired enzymes [126, 138]. For example, to improve the thermostability of XLNs, Song, Tsang, Sylvestre [139] used computational analysis to guide the design of a mutagenesis library containing identified thermal important residues in targeted regions, followed by several rounds of iterative saturation mutagenesis, resulting in an excellent mutant (Xyn10A_{ASPNG}). The thermal inactivation half-life of the obtained mutants was 30-fold longer at 60°C compared to the wild type enzyme [139]. For a similar purpose, Denisenko, Gusakov, Rozhkova, et al. [140] created some mutants of *P. canescens* XLN (*PcXylA*) based on the alignment of amino acids from *PcXylA* with that from a few GH10 XLNs having higher thermostability, and characterized the activity of these mutants [140]. Biochemical characterization showed that one mutant (L18F) displayed increased thermostability with a longer half-life time than wild type enzymes at 50–60°C. Among the different approaches of protein engineering, semi-rational design is regarded as one of the most effective ones, as it combines the benefits of rational design and directed evolution to engineer enzyme activity [126].

The above-mentioned strategies have proven to be useful for the discovery of new enzymes. However, this thesis mainly focusses on the strategy of genome mining.

Various applications of enzymes in industry

Apart from their ability to efficiently degrade xylan in lignocellulosic biomass, xylanolytic enzymes have been reported to play considerable roles in biotechnological processes for various industrial applications [35, 141, 142]. Here, we provide an overview of their application in the following fields: 1) food industry, 2) animal feed, 3) biofuel production, 4) pulp and paper industry, 5) medical and pharmacological industry.

Food industry

Enzymatic processing of food has distinct advantages over traditional chemical-based technology, in that there is less waste and by-products, less energy consumption, and decreased environmental impact. Fungal enzymes are utilized in food industry for modification of products, rather than their full degradation. The most widely used xylanolytic enzymes is XLNs, which have been employed as additive in different baking products, e.g., bread and biscuits. XLNs, as dough strengtheners, are alternatives to chemical dough conditioners, because of their excellent tolerance to the dough with respect to variations in processing parameters and in flour quality [143]. They can make the doughs soft and slack in rye baking, and improve the quality of whole wheat bread, by increasing the specific bread volume [144]. In biscuits, addition of XLNs can improve the texture and taste [145]. In addition, the product of enzymatic lignocellulose degradation can also be used as additives in food industry. Ferulic acid is a released product from AX by FAEs, which has been regarded as

precursor for the synthesis of flavor compounds, such as vanillin and guaiacol. These are aroma compounds used in foods, beverages and fragrance industries [146-148]. Similarly, D-xylose is the hydrolysis product of xylans by BXLs, which can be further used for xylitol production [149-151]. Xylitol is a polyalcohol, which can be employed as a natural food sweetener, dental caries reducer and a sugar substitute for diabetics [149, 151].

Animal feed

Feed ingredients used in monogastric diets (i.e., pigs and poultry) usually contain cellulose and hemicellulose, which are indigestible as these components or fractions have no match with the animals' endogenous enzymes [152]. Some fractions, such as β -glucans, pentosans, and phytates are even considered antinutritive [153]. To ensure maximum uptake efficiency, the supplementation of exogenous enzymes in animal diets has been widely accepted [154]. Xylanolytic enzymes are commonly added to diets containing cereals such as wheat, barley, and rye. These diets contain large amount of water-insoluble AX, which is considered to be the major limiting factor for use of cereals in animal feed [155, 156]. XLNs can hydrolyze water-insoluble AX to its soluble form by breaking down the backbone of the polysaccharide, thereby reducing viscosity in the diet and improving digestibility by the animal [157]. The arabino-xylooligosaccharides as (by-)product from the hydrolysis of AX by XLNs, have been reported to exert positive effects and may act as prebiotics [158]. Apart from monogastric diets, xylanolytic enzymes are also commonly used in ruminant diets, to potentially improve fiber degradability and subsequently increase the intake of digestible energy [159]. For example, FAEs can hydrolyze the linkages between lignin and the cell wall polysaccharides of forages, and therefore increase the accessibility for other enzymes (i.e., cellulases, XLNs) to degrade and improve the nutritive value [160-162]. Recently, related FAE products for feed have become commercially available [163].

Biofuel production

Fossil fuels are the most widely used energy sources worldwide, which are also the principal carbon dioxide emission source [164]. Given fossil-derived fuels are nonrenewable, their depletion has been identified as a future challenge. Thus, biofuels derived from plant biomass, as a sustainable and clean alternative to fossil fuels, are attracting growing attention [165]. Biofuels can be converted and produced from lignocellulosic biomass, while one of the major limitations for efficient conversion is the recalcitrant nature of the plant cell wall [166]. Plant cell wall mainly consists of cellulose, hemicellulose and lignin, of which cellulose and hemicellulose are expected to be significant substrates in the future for bioconversion to biomethanol, bioethanol, or other higher molecular weight alcohols [167]. Xylan represents the most abundant hemicellulose. Xylanolytic enzymes can hydrolyze xylan into its constituent sugars, mainly D-xylose and L-arabinose, for subsequently fermented by microorganisms to produce biofuels. Based on xylan structure, to efficiently depolymerize xylan to the component monosaccharides, a mixture of different enzymatic functionalities is required, including XLNs, BXLs, ABFs, AGUs, FAEs, and AXEs [168, 169]. Additionally, xylanolytic enzymes (especially XLNs) in combination with cellulases/laccases/amylases have been reported to be greatly benefit bioethanol production, which significantly increase the ethanol yield [170, 171].

Pulp and paper industry

Traditionally, the bleaching process used within the pulp and paper industry is divided into three stages, which involves the utilization of different chemicals in each step, such as chlorine dioxide in the first stage, sodium hydroxide and hydrogen peroxide in the second stage, and chlorine peroxides in the final stage. The obvious drawbacks associated with traditional process are the high chemical loads, harmful gas emission rates and the high economic costs. Xylanolytic enzymes as environmentally friendly alternatives have been employed in pulp and paper to minimize these problems [172]. For example, XLNs can partially interrupt the bonds of lignin-carbohydrate complexes by hydrolysis of xylan, and thereby increase the accessibility of the subsequent bleaching chemicals to the pulps, as

well as facilitate the removal of lignin. This action reduces the consumption of chlorine and chlorine dioxide, and significantly improves the final brightness value of pulp [172-175]. Additionally, other xylanolytic enzymes, like AGUs and FAEs, are also reported to be used for pulp and paper industry. Paper and pulp sectors are the traditional applications of softwood, which contain around 5-10% GAX. AGU can remove (Me)GlcA residues of xylan, allowing XLNs to access the xylan backbone, thereby improving the efficiency of xylan degradation [176]. Annual plant fibers are also a cheap alternative source of raw materials for pulp and paper production, which contain up to 35% GAX. The combination of FAEs with other enzymes can promote xylan degradation, thus further improving pulp bleaching processes [177].

Medical and pharmacological industry

Some by-products of xylanolytic enzymes show pharmaceutical and health beneficial functions, such as ferulic acid and other hydroxycinnamic acids from the hydrolysis of xylan by FAEs. They show antimicrobial, anti-inflammatory, anti-diabetic, anti-thrombosis, anti-cancer, and cholesterol-lowering agents, which are of interest for a wide range of applications in medical and pharmaceutical industries [81, 178, 179]. (Arabino)xylo-oligosaccharides are products of xylan degradation, released during XLN treatment, which possess prebiotic effects. They can stimulate the growth of beneficial bacteria in the human gut, such as *Bifidobacteria*, while restricting harmful bacteria [180].

Other applications

Apart from the above-mentioned applications, xylanolytic enzymes can also be used in other applications. For example, in detergents, an FAE-containing multi-enzyme system is used to improve the performance of liquid laundry detergents particularly at low temperature [181], while XLN-containing detergents can effectively remove plant stains [182]. In the beverage industry, adding XLNs and ABFs in juice clarification and extraction can result in improved yield, stabilization of fruit pulp, reduction of viscosity, and clearing of the juice [94, 183]. The addition of ABFs also can prevent haze formation in fruit juice [183]. In seasonings and alcoholic beverages, FAEs have been used for both removing off-flavors/odors as well as enhancing the aroma [81]. Ferulic acid produced by FAEs is also used as additive in beer as it is one of the major antioxidant constituents [184].

Objective and outline of the thesis

With the development of the bio-based economy, industrial catalysis is increasingly dependent on enzymes. Although the industrial applications mentioned above already account for over 80% of the global market of enzymes, there is still scope for further development of the market as new enzymes and enzymatic processes emerge [185]. In the genome era, the explosive growth of fungal genomes has led to the discovery of an exponentially increasing number of CAZyme-encoding genes, yet most of them lack functional annotations [104, 122]. Additionally, fungal genome comparison pointed out that some fungi have clear expansions in specific CAZy families (**Table 2**). These expansions are expected to widen the range of substrates that can be degraded by the fungus or provide it with enzymes with different properties. Since biochemical characterization has not been able to keep up with this growth of new putative CAZymes, many CAZy families have characterized fungal enzymes accounting for less than 20% of the total enzyme count among the CAZy families associated with plant biomass degradation [104, 122]. This lack of insight into many CAZy families in turn prevents reliable functional annotation of fungal genomes. Therefore, selection of CAZyme-encoding genes with unknown function from fungal genomes, heterologous expression, and characterization of the corresponding proteins could not only discover new enzymes with industrially desired properties but also improve our understanding of the poorly characterized CAZy families. The aim of this thesis is to selectively characterize fungal enzymes with unknown functions from the expanded and poorly characterized CAZy families associated with hemicellulose (mainly xylan) degradation through a genome mining combined with functional characterization strategies. Results described in

this thesis will not only provide the CAZyme community with new enzymes and expand the toolbox of enzymes for bioprocessing, but will also improve the functional annotations of fungal genomes and deepen our understanding of certain aspects of fungal physiology. Below is a brief description of the content of the chapters in this thesis.

In **Chapter 2**, we used a genome-mining strategy to add 87 additional CE1 sequences from 18 genome sequenced fungal species into the previously established CE1 phylogenetic tree, having five subfamilies (CE1_SF1–SF5). To explore the differences between SF1 and SF2, six ascomycete candidates from the unknown branches of SF1 and SF2 were selected and characterized. The results showed the enzymes from CE1_SF1 only exhibited AXE activity, whereas an enzyme from CE1_SF2 possessed dual FAE/AXE activity. This dual activity enzyme supports the evolutionary relationship of CE1_SF1 and SF2. Enzymes in this chapter can be used as an additional ingredient for enzyme cocktails to remove ester-linked decorations, which enables access for the backbone degrading enzymes.

A similar strategy was applied in **Chapter 3**. Using a genome-mining strategy through a BLAST search against 38 genome sequenced fungal species, we added over 100 amino acid sequences to the set of characterized GH30 enzymes in CAZy and uncharacterized bacterial and animal enzymes in NCBI database, and constructed a GH30 phylogenetic tree. Eleven candidates were selected from the various fungal branches for biochemical characterization. Our results confirmed the diverse activities in the GH30 fungal subfamilies (e.g., β -(1→6)-glucanase, β -(1→6)-galactobiohydrolase, β -(1→4)-xylobiohydrolase, endoxylanase). Additionally, combining characterization data and phylogenetic analysis, we proposed a new fungal subfamily in GH30, GH30_11, which displays β -(1→6)-galactobiohydrolase. This study demonstrated that fungal genome mining is not limited to the discovery of new enzymes, but can also establish new fungal subfamilies. Enzymes discovered in this Chapter could be used to produce different oligosaccharides (especially disaccharides) with high prebiotic potential.

The ability to discover new fungal enzymes from poorly characterized families by genome mining approach was well demonstrated in the studies of **Chapter 2** and **3**, while there is also interest in new enzymes from expanded families containing plant biomass degrading enzymes. Previously, a comparison of gene numbers in different fungal genomes of CAZymes related to plant biomass degradation revealed that *P. subrubescens* showed strong expansions in specific CAZy families and/or certain activities [104]. To determine whether this expansion resulted in functional diversification of these enzymes or led to redundancy, in **Chapter 4** and **Chapter 5**, we characterized multiple *P. subrubescens* ABFs and XLNs from different families, respectively.

In **Chapter 4**, we focused on the characterization of nine *P. subrubescens* ABFs from GH51, GH54 and GH62. The results showed that *P. subrubescens* ABFs (*Ps*ABFs) had highly diverse specificity and activity levels, indicating that the expansion of *Ps*ABFs was accompanied by diversification of the enzymes. That the expansion of CAZymes in the *P. subrubescens* genome is accompanied by functional diversity was also verified in *P. subrubescens* XLNs from GH10 and GH11, which are presented in **Chapter 5**. *P. subrubescens* XLNs (*Ps*XLNs) had distinct product profiles during wheat arabinoxylan (WAX) hydrolysis and diverse biochemical properties. In addition, it is worth mentioning that *Ps*XLNs in **Chapter 5** showed significant cooperative action with *Ps*ABFs from GH51 or GH54 in **Chapter 4** for WAX degradation. This finding may reflect an evolutionary adaptation of this species that provides a wider enzymatic toolbox for synergistic degradation of hemicellulose in its natural habitat, which also gives insight into a more diverse XLN and ABF system for the efficient degradation of complex hemicelluloses.

The enzymes described in **Chapter 2-5** are involved in the degradation of xylan. Xyloglucan is another important hemicellulose component in addition to xylan, and its degradation is essential for overall lignocellulosic biomass utilization. In **Chapter 6**, we described the discovery and characterization of new fungal xyloglucanases (XEGs) in GH44. So far, GH44 has been shown to contain bacterial XEGs only. Genome analysis revealed GH44 members in fungal species from the phylum Basidiomycota, but not in other fungi. To determine whether these members are XEGs,

two basidiomycete GH44 enzymes were selected and characterized. The results showed that these two enzymes exhibited XEG activity and displayed a hydrolytic cleavage pattern different from that observed in fungal XEGs from other GH families. Fungal GH44 XEGs provided in this study represent a new class of enzymes for plant biomass conversion and valorization.

The results are summarized and discussed in **Chapter 7**.

References

1. Ho, D.P., H.H. Ngo, and W. Guo, *A mini review on renewable sources for biofuel*. **Bioresource Technology**, 2014. 169: p. 742-749.
2. Cintas, O., et al., *Geospatial supply-demand modeling of lignocellulosic biomass for electricity and biofuels in the European Union*. **Biomass and Bioenergy**, 2021. 144: p. 105870.
3. Usmani, Z., et al., *Bioprocessing of waste biomass for sustainable product development and minimizing environmental impact*. **Bioresource Technology**, 2020: p. 124548.
4. Iqbal, H.M.N., G. Kyazze, and T. Keshavarz, *Advances in the valorization of lignocellulosic materials by biotechnology: an overview*. **BioResources**, 2013. 8(2): p. 3157-3176.
5. Xu, F., et al., *Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: a mini-review*. **Applied Energy**, 2013. 104: p. 801-809.
6. Krässig, H.A., *Cellulose: structure, accessibility and reactivity*. **Polymer Monographs**, 1993. 11: p. 240.
7. Wang, S., et al., *Pyrolysis behaviors of four O-acetyl-preserved hemicelluloses isolated from hardwoods and softwoods*. **Fuel**, 2015. 150: p. 243-251.
8. Scheller, H.V. and P. Ulvskov, *Hemicelluloses*. **Annual Review of Plant Biology**, 2010. 61: p. 263-289.
9. Ebringerová, A. and T. Heinze, *Xylan and xylan derivatives—biopolymers with valuable properties, I. Naturally occurring xylns structures, isolation procedures and properties*. **Macromolecular Rapid Communications**, 2000. 21(9): p. 542-556.
10. Stefanidis, S.D., et al., *A study of lignocellulosic biomass pyrolysis via the pyrolysis of cellulose, hemicellulose and lignin*. **Journal of Analytical and Applied Pyrolysis**, 2014. 105: p. 143-150.
11. Lange, L., et al., *Developing a sustainable and circular bio-based economy in EU: by partnering across sectors, upscaling and using new knowledge faster, and for the benefit of climate, environment & biodiversity, and people & business*. **Frontiers in Bioengineering and Biotechnology**, 2021: p. 1456.
12. Lange, L., V. Parmar, and A. Meyer, *Biocatalysis*. **Encyclopedia of Sustainable Technologies**, 2017: p. 663-673.
13. Abraham, M., *Encyclopedia of sustainable technologies*. **Elsevier Science**, 2017.
14. Doblin, M.S., F. Pettolino, and A. Bacic, *Plant cell walls: the skeleton of the plant world*. **Functional Plant Biology**, 2010. 37(5): p. 357-381.
15. Rytioja, J., et al., *Plant-polysaccharide-degrading enzymes from basidiomycetes*. **Microbiology and Molecular Biology Reviews**, 2014. 78(4): p. 614-649.
16. Chanda, S., et al., 262. *The constitution of xylan from esparto grass (stipa tenacissima, L.)*. **Journal of the Chemical Society (Resumed)**, 1950: p. 1289-1297.
17. Montgomery, R., F. Smith, and H. Srivastava, *Structure of Corn Hull Hemicellulose. I. Partial Hydrolysis and Identification of 2-O-(α -D-Glucopyranosyluronic Acid)-D-xylopyranose 1, 2*. **Journal of the American Chemical Society**, 1956. 78(12): p. 2837-2839.
18. Eda, S., A. Ohnishi, and K. Katō, *Xylan isolated from the stalk of Nicotiana tabacum*. **Agricultural and Biological Chemistry**, 1976. 40(2): p. 359-364.
19. Deniaud, E., et al., *Structural studies of the mix-linked β -(1 \rightarrow 3)/ β -(1 \rightarrow 4)-D-xylns from the cell wall of *Palmaria palmata* (Rhodophyta)*. **International Journal of Biological Macromolecules**, 2003. 33(1-3): p. 9-18.
20. Girio, F.M., et al., *Hemicelluloses for fuel ethanol: a review*. **Bioresource Technology**, 2010. 101(13): p. 4775-4800.
21. Vuong, T.V. and E.R. Master, *Enzymatic upgrading of heteroxylns for added-value chemicals and polymers*. **Current Opinion in Biotechnology**, 2022. 73: p. 51-60.
22. Beg, Q., et al., *Microbial xylanases and their industrial applications: a review*. **Applied Microbiology and Biotechnology**, 2001. 56(3): p. 326-338.
23. Hatfield, R.D., D.M. Rancour, and J.M. Marita, *Grass cell walls: a story of cross-linking*. **Frontiers in Plant**

- Science*, 2017. 7: p. 2056.
24. de Carvalho, D.M., et al., *Improving the thermal stability of different types of xylan by acetylation. Carbohydrate Polymers*, 2019. 220: p. 132-140.
 25. Smith, P.J., et al., *Designer biomass for next-generation biorefineries: leveraging recent insights into xylan structure and biosynthesis. Biotechnology for Biofuels*, 2017. 10(1): p. 1-14.
 26. Gao, Y., et al., *A grass-specific cellulose-xylan interaction dominates in sorghum secondary cell walls. Nature Communications*, 2020. 11(1): p. 1-10.
 27. DORCZYK, M.S.I. and C.G. BILIADERIS, *Structural Heterogeneity of Wheat Endosperm Arabinoxylans'. Cereal Chemistry*, 1993. 70(6): p. 641-646.
 28. Biely, P., S. Singh, and V. Puchart, *Towards enzymatic breakdown of complex plant xylan structures: state of the art. Biotechnology Advances*, 2016. 34(7): p. 1260-1274.
 29. Madhavan, A., A. Pandey, and R.K. Sukumaran, *Expression system for heterologous protein expression in the filamentous fungus Aspergillus unguis. Bioresource Technology*, 2017. 245: p. 1334-1342.
 30. Nevalainen, K.H., V.S. Te'o, and P.L. Bergquist, *Heterologous protein expression in filamentous fungi. Trends in Biotechnology*, 2005. 23(9): p. 468-474.
 31. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research*, 2014. 42(D1): p. D490-D495.
 32. Simmons, T.J., et al., *Structural and electronic determinants of lytic polysaccharide monoxygenase reactivity on polysaccharide substrates. Nature Communications*, 2017. 8(1): p. 1-12.
 33. Basotra, N., et al., *Characterization of a novel lytic polysaccharide monoxygenase from Malbranchea cinnamomea exhibiting dual catalytic behavior. Carbohydrate Research*, 2019. 478: p. 46-53.
 34. Ipsen, J.Ø., et al., *Lytic polysaccharide monoxygenases and other histidine-brace copper proteins: structure, oxygen activation and biotechnological applications. Biochemical Society Transactions*, 2021. 49(1): p. 531-540.
 35. Lian, Z., et al., *An integrated process to produce prebiotic xylooligosaccharides by autohydrolysis, nano-filtration and endo-xylanase from alkali-extracted xylan. Bioresource Technology*, 2020. 314: p. 123685.
 36. Agrawal, D., A. Tsang, and B.S. Chadha, *Economizing the lignocellulosic hydrolysis process using heterologously expressed auxiliary enzymes feruloyl esterase D (CE1) and β -xylosidase (GH43) derived from thermophilic fungi Scytalidium thermophilum. Bioresource Technology*, 2021. 339: p. 125603.
 37. Šuchová, K., et al., *A novel GH30 xylobiohydrolase from Acremonium alcalophilum releasing xylobiose from the non-reducing end. Enzyme and Microbial Technology*, 2020. 134: p. 109484.
 38. Paës, G., J.-G. Berrin, and J. Beaugrand, *GH11 xylanases: structure/function/properties relationships and applications. Biotechnology Advances*, 2012. 30(3): p. 564-592.
 39. Pollet, A., J.A. Delcour, and C.M. Courtin, *Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families. Critical Reviews in Biotechnology*, 2010. 30(3): p. 176-191.
 40. Knob, A., C.F. Terrasan, and E. Carmona, *β -Xylosidases from filamentous fungi: an overview. World Journal of Microbiology and Biotechnology*, 2010. 26(3): p. 389-407.
 41. Bosetto, A., et al., *Research progress concerning fungal and bacterial β -xylosidases. Applied Biochemistry and Biotechnology*, 2016. 178(4): p. 766-795.
 42. Bhardwaj, N., B. Kumar, and P. Verma, *A detailed overview of xylanases: an emerging biomolecule for current and future prospective. Bioresources and Bioprocessing*, 2019. 6(1): p. 1-36.
 43. Ravanal, M.C., E. Callegari, and J. Eyzaguirre, *Novel bifunctional α -L-arabinofuranosidase/xylobiohydrolase (ABF3) from Penicillium purpurogenum. Applied and Environmental Microbiology*, 2010. 76(15): p. 5247-5253.
 44. Katsimpouras, C., et al., *A novel fungal GH30 xylanase with xylobiohydrolase auxiliary activity. Biotechnology for Biofuels*, 2019. 12(1): p. 120.
 45. Nakamichi, Y., et al., *Structural and functional characterization of a bifunctional GH30-7 xylanase B from the filamentous fungus Talaromyces cellulolyticus. Journal of Biological Chemistry*, 2019. 294(11): p. 4065-4078.
 46. Li, X., et al., *Glycoside Hydrolase family 30 harbors fungal subfamilies with distinct polysaccharide specificities. New Biotechnology*, 2022. 67: p. 32-41.
 47. Chadha, B.S., et al., *Thermostable xylanases from thermophilic fungi and bacteria: current perspective. Bioresource Technology*, 2019. 277: p. 195-203.
 48. Kolenová, K., M. Vršanská, and P. Biely, *Mode of action of endo- β -1, 4-xylanases of families 10 and 11 on acidic xylooligosaccharides. Journal of Biotechnology*, 2006. 121(3): p. 338-345.

49. Bauer, S., et al., *Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls. Proceedings of the National Academy of Sciences*, 2006. 103(30): p. 11417-11422.
50. Wakiyama, M., et al., *Purification and properties of an extracellular β -xylosidase from *Aspergillus japonicus* and sequence analysis of the encoding gene. Journal of Bioscience and Bioengineering*, 2008. 106(4): p. 398-404.
51. Kitamoto, N., et al., *Sequence analysis, overexpression, and antisense inhibition of a β -xylosidase gene, *xylA*, from *Aspergillus oryzae* KBN616. Applied and Environmental Microbiology*, 1999. 65(1): p. 20-24.
52. Yang, X., et al., *Two xylose-tolerant GH43 bifunctional β -xylosidase/ α -arabinosidases and one GH11 xyylanase from *Humicola insolens* and their synergy in the degradation of xylan. Food Chemistry*, 2014. 148: p. 381-387.
53. Jordan, D.B., et al., *Structure-function relationships of a catalytically efficient β -D-xylosidase. Applied Biochemistry and Biotechnology*, 2007. 141(1): p. 51-76.
54. Herrmann, M.C., et al., *The β -D-xylosidase of *Trichoderma reesei* is a multifunctional β -D-xylan xylohydrolase. Biochemical Journal*, 1997. 321(2): p. 375-381.
55. Kurakake, M., et al., *Characteristics of transxylosylation by β -xylosidase from *Aspergillus awamori* K4. Biochimica et Biophysica Acta (BBA)-General Subjects*, 2005. 1726(3): p. 272-279.
56. Sakamoto, T. and H. Kawasaki, *Purification and properties of two type-B α -L-arabinofuranosidases produced by *Penicillium chrysogenum*. Biochimica et Biophysica Acta (BBA)-General Subjects*, 2003. 1621(2): p. 204-210.
57. Kaur, A.P., et al., *Functional and structural diversity in GH 62 α -L-arabinofuranosidases from the thermophilic fungus *S cyathidium thermophilum*. Microbial Biotechnology*, 2015. 8(3): p. 419-433.
58. Pouvreau, L., et al., *Chryso sporium lucknowense C1 arabinofuranosidases are selective in releasing arabinose from either single or double substituted xylose residues in arabinoxylans. Enzyme and Microbial Technology*, 2011. 48(4-5): p. 397-403.
59. Beldman, G., et al., *Arabinans and Arabinan Degrading Enzymes. Advances in Macromolecular Carbohydrate Research*, 1997: p. 1-64.
60. Maehara, T., et al., *Crystal structure and characterization of the glycoside hydrolase family 62 α -L-arabinofuranosidase from *Streptomyces coelicolor*. Journal of Biological Chemistry*, 2014. 289(11): p. 7962-7972.
61. Ohta, K., S. Fujii, and C. Higashida, *Characterization of a glycoside hydrolase family-51 α -l-arabinofuranosidase gene from *Aureobasidium pullulans* ATCC 20524 and its encoded product. Journal of Bioscience and Bioengineering*, 2013. 116(3): p. 287-292.
62. Contesini, F.J., et al., *Structural and functional characterization of a highly secreted α -L-arabinofuranosidase (GH62) from *Aspergillus nidulans* grown on sugarcane bagasse. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 2017. 1865(12): p. 1758-1769.
63. Siguier, B., et al., *First structural insights into α -L-arabinofuranosidases from the two GH62 glycoside hydrolase subfamilies. Journal of Biological Chemistry*, 2014. 289(8): p. 5261-5273.
64. Gielkens, M.M., J. Visser, and L.H. de Graaff, *Arabinoxylan degradation by fungi: characterization of the arabinoxylan-arabinofuranohydrolase encoding genes from *Aspergillus niger* and *Aspergillus tubingensis*. Current Genetics*, 1997. 31(1): p. 22-29.
65. Sakamoto, T., et al., *Identification of a GH62 α -L-arabinofuranosidase specific for arabinoxylan produced by *Penicillium chrysogenum*. Applied Microbiology and Biotechnology*, 2011. 90(1): p. 137-146.
66. Kormelink, F., et al., *Purification and characterization of a (1, 4)- β -D-arabinoxylan arabinofuranohydrolase from *Aspergillus awamori*. Applied Microbiology and Biotechnology*, 1991. 35(6): p. 753-758.
67. Tenkanen, M. and M. Siika-aho, *An α -glucuronidase of *Schizophyllum commune* acting on polymeric xylan. Journal of Biotechnology*, 2000. 78(2): p. 149-161.
68. Ryabova, O., et al., *A novel family of hemicellulolytic α -glucuronidase. FEBS Letters*, 2009. 583(9): p. 1457-1462.
69. Martínez, P.M., et al., *The two *Rasamsonia emersonii* α -glucuronidases, Re GH67 and Re GH115, show a different mode-of-action towards glucuronoxylan and glucuronoxyloligosaccharides. Biotechnology for Biofuels*, 2016. 9(1): p. 1-10.
70. Chong, S.-L., et al., *The α -glucuronidase *Agu1* from *Schizophyllum commune* is a member of a novel glycoside hydrolase family (GH115). Applied Microbiology and Biotechnology*, 2011. 90(4): p. 1323-1332.
71. de Vries, R.P., et al., *aguA, the gene encoding an extracellular α -glucuronidase from *Aspergillus tubingensis*, is specifically induced on xylose and not on glucuronic acid. Journal of Bacteriology*, 1998. 180(2):

- p. 243-249.
72. Siika-aho, M., et al., *An α -glucuronidase from Trichoderma reesei RUT C-30. Enzyme and Microbial Technology*, 1994. 16(9): p. 813-819.
 73. Biely, P., et al., *Inverting character of α -glucuronidase A from Aspergillus tubingensis. Biochimica et Biophysica Acta (BBA)-General Subjects*, 2000. 1474(3): p. 360-364.
 74. Altaner, C., et al., *Regioselective deacetylation of cellulose acetates by acetyl xylan esterases of different CE-families. Journal of Biotechnology*, 2003. 105(1-2): p. 95-104.
 75. Karnaouri, A., et al., *Thermophilic enzyme systems for efficient conversion of lignocellulose to valuable products: Structural insights and future perspectives for esterases and oxidative catalysts. Bioresource Technology*, 2019. 279: p. 362-372.
 76. Uhliariková, I., et al., *Positional specificity of acetyl-xylan esterases on natural polysaccharide: An NMR study. Biochimica et Biophysica Acta (BBA)-General Subjects*, 2013. 1830(6): p. 3365-3372.
 77. Mai-Gisondi, G., et al., *Functional comparison of versatile carbohydrate esterases from families CE1, CE6 and CE16 on acetyl-4-O-methylglucuronoxylan and acetyl-galactoglucomannan. Biochimica et Biophysica Acta (BBA)-General Subjects*, 2017. 1861(9): p. 2398-2405.
 78. Biely, P., et al., *Trichoderma reesei CE16 acetyl esterase and its role in enzymatic degradation of acetylated hemicellulose. Biochimica et Biophysica Acta (BBA)-General Subjects*, 2014. 1840(1): p. 516-525.
 79. Puchart, V., et al., *Comparison of fungal carbohydrate esterases of family CE16 on artificial and natural substrates. Journal of Biotechnology*, 2016. 233: p. 228-236.
 80. Crepin, V.F., C.B. Faulds, and I. Connerton, *Functional classification of the microbial feruloyl esterases. Applied Microbiology and Biotechnology*, 2004. 63(6): p. 647-652.
 81. Dilokpimol, A., et al., *Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. Biotechnology for Biofuels*, 2016. 9(1): p. 1-18.
 82. Benoit, I., et al., *Biotechnological applications and potential of fungal feruloyl esterases based on prevalence, classification and biochemical diversity. Biotechnology Letters*, 2008. 30(3): p. 387-396.
 83. Špániková, S. and P. Biely, *Glucuronoyl esterase—novel carbohydrate esterase produced by Schizophyllum commune. FEBS Letters*, 2006. 580(19): p. 4597-4601.
 84. Bååth, J.A., et al., *Biochemical and structural features of diverse bacterial glucuronoyl esterases facilitating recalcitrant biomass conversion. Biotechnology for Biofuels*, 2018. 11(1): p. 1-14.
 85. Mosbech, C., et al., *The natural catalytic function of CuGE glucuronoyl esterase in hydrolysis of genuine lignin-carbohydrate complexes from birch. Biotechnology for Biofuels*, 2018. 11(1): p. 1-9.
 86. Hüttner, S., et al., *Specific xylan activity revealed for AA9 lytic polysaccharide monoxygenases of the thermophilic fungus Malbranchea cinnamomea by functional characterization. Applied and Environmental Microbiology*, 2019. 85(23).
 87. Frommhagen, M., et al., *Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monoxygenase. Biotechnology for Biofuels*, 2015. 8(1): p. 1-12.
 88. Couturier, M., et al., *Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. Nature Chemical Biology*, 2018. 14(3): p. 306.
 89. Quinlan, R.J., et al., *Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. Proceedings of the National Academy of Sciences*, 2011. 108(37): p. 15079-15084.
 90. Zerva, A., et al., *A new synergistic relationship between xylan-active LPMO and xylobiohydrolase to tackle recalcitrant xylan. Biotechnology for Biofuels*, 2020. 13(1): p. 1-13.
 91. Malgas, S., et al., *A mini review of xylanolytic enzymes with regards to their synergistic interactions during hetero-xylan degradation. World Journal of Microbiology and Biotechnology*, 2019. 35(12): p. 187.
 92. De Vries, R.P., et al., *Synergy between enzymes from Aspergillus involved in the degradation of plant cell wall polysaccharides. Carbohydrate Research*, 2000. 327(4): p. 401-410.
 93. Bennett, J., *Mycotechnology: the role of fungi in biotechnology. Journal of Biotechnology*, 1998. 66(2-3): p. 101-107.
 94. Polizeli, M., et al., *Xylanases from fungi: properties and industrial applications. Applied Microbiology and Biotechnology*, 2005. 67(5): p. 577-591.
 95. Hawksworth, D.L. and R. Lücking, *Fungal diversity revisited: 2.2 to 3.8 million species. The Fungal Kingdom*, 2017: p. 79-95.
 96. Ogawa, J. and S. Shimizu, *Microbial enzymes: new industrial applications from traditional screening methods. Trends in Biotechnology*, 1999. 17(1): p. 13-20.

97. Ma, J., L. Wu, and I. Kurtböke, *Skerman and beyond: 2019 status of the Global Catalogue of Microorganisms*. **Microbiology Australia**, 2019. 40(3): p. 121-124.
98. Pedersen, M., et al., *Identification of thermostable β -xylosidase activities produced by *Aspergillus brasiliensis* and *Aspergillus niger**. **Biotechnology Letters**, 2007. 29(5): p. 743-748.
99. Ang, S., et al., *Isolation, screening, and identification of potential cellulolytic and xylanolytic producers for biodegradation of untreated oil palm trunk and its application in saccharification of lemongrass leaves*. **Preparative Biochemistry and Biotechnology**, 2015. 45(3): p. 279-305.
100. Krogh, K.B., et al. *Screening genus *Penicillium* for producers of cellulolytic and xylanolytic enzymes*. **Applied Biochemistry and Biotechnology**, 2004. 114: p. 308-401.
101. Ja'afaru, M.I., *Screening of fungi isolated from environmental samples for xylanase and cellulase production*. **International Scholarly Research Notices**, 2013. 2013.
102. Sridevi, B. and M.S. Charya, *Isolation, identification and screening of potential cellulase-free xylanase producing fungi*. **African Journal of Biotechnology**, 2011. 10(22): p. 4624-4630.
103. Ferrer, M., F. Martínez-Abarca, and P.N. Golyshein, *Mining genomes and 'metagenomes' for novel catalysts*. **Current Opinion in Biotechnology**, 2005. 16(6): p. 588-593.
104. de Vries, R.P. and M.R. Mäkelä, *Genomic and Postgenomic Diversity of Fungal Plant Biomass Degradation Approaches*. **Trends in Microbiology**, 2020. 28(6): p. 487-499.
105. Lämmle, K., et al., *Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning*. **Journal of Biotechnology**, 2007. 127(4): p. 575-592.
106. Schloss, P.D. and J. Handelsman, *Biotechnological prospects from metagenomics*. **Current Opinion in Biotechnology**, 2003. 14(3): p. 303-310.
107. Daniel, R., *The soil metagenome—a rich resource for the discovery of novel natural products*. **Current Opinion in Biotechnology**, 2004. 15(3): p. 199-204.
108. Uchiyama, T. and K. Miyazaki, *Functional metagenomics for enzyme discovery: challenges to efficient screening*. **Current Opinion in Biotechnology**, 2009. 20(6): p. 616-622.
109. Lorenz, P. and C. Schleper, *Metagenome—a challenging source of enzyme discovery*. **Journal of Molecular Catalysis B: Enzymatic**, 2002. 19: p. 13-19.
110. Kim, Y.-J., et al., *Screening and characterization of a novel esterase from a metagenomic library*. **Protein Expression and Purification**, 2006. 45(2): p. 315-323.
111. Kwon, E.J., et al., *Construction of a metagenomic library from compost and screening of cellulase-and xylanase-positive clones*. **Journal of the Korean Society for Applied Biological Chemistry**, 2010. 53(6): p. 702-708.
112. Ferrer, M., et al., *Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora*. **Environmental Microbiology**, 2005. 7(12): p. 1996-2010.
113. Steele, H.L., et al., *Advances in recovery of novel biocatalysts from metagenomes*. **Journal of Molecular Microbiology and Biotechnology**, 2009. 16(1-2): p. 25-37.
114. Matsuzawa, T., S. Kaneko, and K. Yaoi, *Screening, identification, and characterization of a GH43 family β -xylosidase/ α -arabinofuranosidase from a compost microbial metagenome*. **Applied Microbiology and Biotechnology**, 2015. 99(21): p. 8943-8954.
115. Li, L.-L., et al., *Bioprospecting metagenomics of decaying wood: mining for new glycoside hydrolases*. **Biotechnology for Biofuels**, 2011. 4(1): p. 1-13.
116. Hu, Y., et al., *Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach*. **Applied Microbiology and Biotechnology**, 2008. 80(5): p. 823-830.
117. Grigoriev, I.V., et al., *MycCosm portal: gearing up for 1000 fungal genomes*. **Nucleic Acids Research**, 2014. 42(D1): p. D699-D704.
118. Peng, M., et al., *The draft genome sequence of the ascomycete fungus *Penicillium subrubescens* reveals a highly enriched content of plant biomass related CAZymes compared to related fungi*. **Journal of Biotechnology**, 2017. 246: p. 1-3.
119. de Vries, R.P., et al., *Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus**. **Genome Biology**, 2017. 18(1): p. 1-45.
120. Dilokpimol, A., et al., **Penicillium subrubescens* adapts its enzyme production to the composition of plant biomass*. **Bioresource Technology**, 2020: p. 123477.
121. Mäkelä, M.R., et al., **Penicillium subrubescens* is a promising alternative for *Aspergillus niger* in enzymatic plant biomass saccharification*. **New Biotechnology**, 2016. 33(6): p. 834-841.
122. de Vries, R.P., I.V. Grigoriev, and A. Tsang, *Introduction: overview of fungal genomics*. **Fungal Genomics**,

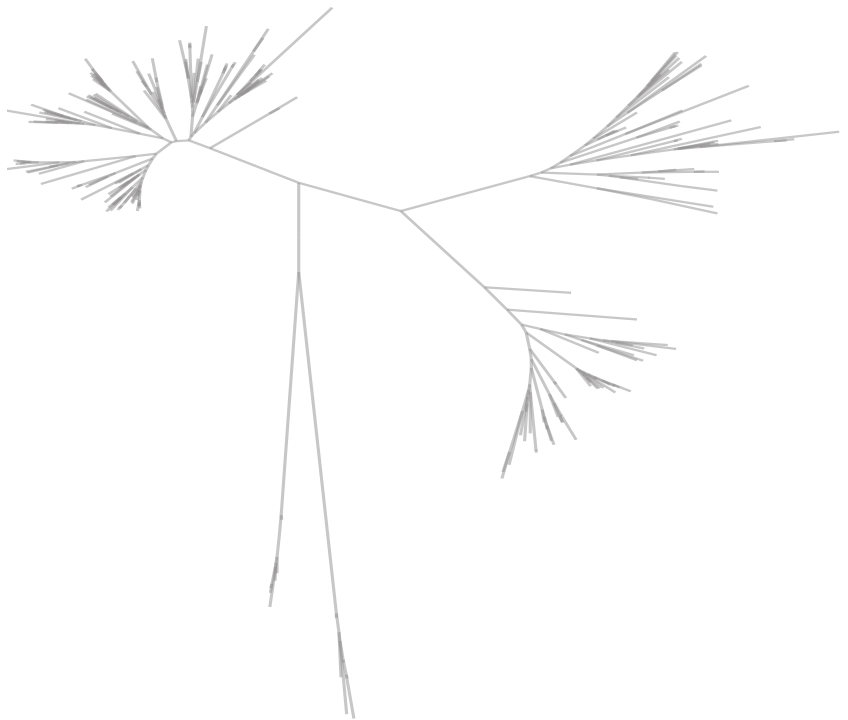
- 2018: p. 1-7.
123. Dilokpimol, A., et al., *Fungal glucuronoyl esterases: genome mining based enzyme discovery and biochemical characterization*. **New Biotechnology**, 2018. 40: p. 282-287.
124. Dilokpimol, A., et al., *Fungal feruloyl esterases: functional validation of genome mining based enzyme discovery including uncharacterized subfamilies*. **New Biotechnology**, 2018. 41: p. 9-14.
125. Lutz, S. and S.M. Iamurri, *Protein engineering: past, present, and future*. **Protein Engineering**, 2018: p. 1-12.
126. Chica, R.A., N. Doucet, and J.N. Pelletier, *Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design*. **Current Opinion in Biotechnology**, 2005. 16(4): p. 378-384.
127. Striprang, R., et al., *Improvement of thermostability of fungal xylanase by using site-directed mutagenesis*. **Journal of Biotechnology**, 2006. 126(4): p. 454-462.
128. Fenel, F., A.-J. Zitting, and A. Kantelinen, *Increased alkali stability in *Trichoderma reesei* endo-1, 4- β -xylanase II by site directed mutagenesis*. **Journal of Biotechnology**, 2006. 121(1): p. 102-107.
129. Nanda, V. and R.L. Koder, *Designing artificial enzymes by intuition and computation*. **Nature Chemistry**, 2010. 2(1): p. 15-24.
130. Damborsky, J. and J. Brezovsky, *Computational tools for designing and engineering biocatalysts*. **Current Opinion in Chemical Biology**, 2009. 13(1): p. 26-34.
131. Smith, B.A. and M.H. Hecht, *Novel proteins: from fold to function*. **Current Opinion in Chemical Biology**, 2011. 15(3): p. 421-426.
132. Steiner, K. and H. Schwab, *Recent advances in rational approaches for enzyme engineering*. **Computational and Structural Biotechnology Journal**, 2012. 2(3): p. e201209010.
133. Fenel, F., et al., *A de novo designed N-terminal disulphide bridge stabilizes the *Trichoderma reesei* endo-1, 4- β -xylanase II*. **Journal of Biotechnology**, 2004. 108(2): p. 137-143.
134. Bloom, J.D. and F.H. Arnold, *In the light of directed evolution: pathways of adaptive protein evolution*. **Proceedings of the National Academy of Sciences**, 2009. 106(Supplement 1): p. 9995-10000.
135. Bornscheuer, U.T., et al., *Directed evolution empowered redesign of natural proteins for the sustainable production of chemicals and pharmaceuticals*. **Angewandte Chemie International Edition**, 2019. 58(1): p. 36-40.
136. Otten, L.G. and W.J. Quax, *Directed evolution: selecting today's biocatalysts*. **Biomolecular Engineering**, 2005. 22(1-3): p. 1-9.
137. bin Abdul Wahab, M.K.H., M.A. bin Jonet, and R.M. Illias, *Thermostability enhancement of xylanase *Aspergillus fumigatus* RT-1*. **Journal of Molecular Catalysis B: Enzymatic**, 2016. 134: p. 154-163.
138. WANG, F.-y. and W.-y. XUE, *Using a new approach to engineer enzyme activity—semi-rational design*. **Applied Chemical Industry**, 2006. 35(8): p. 634-636.
139. Song, L., A. Tsang, and M. Sylvestre, *Engineering a thermostable fungal GH10 xylanase, importance of N-terminal amino acids*. **Biotechnology and Bioengineering**, 2015. 112(6): p. 1081-1091.
140. Denisenko, Y.A., et al., *Site-directed mutagenesis of GH10 xylanase A from *Penicillium canescens* for determining factors affecting the enzyme thermostability*. **International Journal of Biological Macromolecules**, 2017. 104: p. 665-671.
141. Cho, E.J., et al., *Bioconversion of biomass waste into high value chemicals*. **Bioresource Technology**, 2020. 298: p. 122386.
142. Colonia, B.S.O., et al., *Pulp improvement of oil palm empty fruit bunches associated to solid-state biopulping and biobleaching with xylanase and lignin peroxidase cocktail produced by *Aspergillus* sp. LPB-5*. **Bioresource Technology**, 2019. 285: p. 121361.
143. Shahi, N., et al., *Xylanase: a promising enzyme*. **Journal of Chemical and Pharmaceutical Research**, 2016. 8(3): p. 334-339.
144. Shah, A.R., R. Shah, and D. Madamwar, *Improvement of the quality of whole wheat bread by supplementation of xylanase from *Aspergillus foetidus**. **Bioresource Technology**, 2006. 97(16): p. 2047-2053.
145. Raveendran, S., et al., *Applications of microbial enzymes in food industry*. **Food Technology and Biotechnology**, 2018. 56(1): p. 16-30.
146. Di Gioia, D., et al., *Metabolic engineering of *Pseudomonas fluorescens* for the production of vanillin from ferulic acid*. **Journal of Biotechnology**, 2011. 156(4): p. 309-316.
147. Priefert, H., J. Rabenhorst, and A. Steinbüchel, *Biotechnological production of vanillin*. **Applied Microbiology and Biotechnology**, 2001. 56(3): p. 296-314.

148. Gallage, N.J., et al., *Vanillin formation from ferulic acid in Vanilla planifolia is catalysed by a single enzyme*. **Nature Communications**, 2014. 5(1): p. 1-14.
149. Baptista, S.L., et al., *Xylitol production from lignocellulosic whole slurry corn cob by engineered industrial *Saccharomyces cerevisiae* PE-2*. **Bioresource Technology**, 2018. 267: p. 481-491.
150. Romani, A., et al., *Aqueous solutions of deep eutectic systems as reaction media for the saccharification and fermentation of hardwood xylan into xylitol*. **Bioresource Technology**, 2020. 311: p. 123524.
151. Baptista, S., A.P. Romani, and L. Domingues, *Biotechnological advancements, innovations and challenges for sustainable xylitol production by yeast*. **Reference Module in Life Sciences**, 2020.
152. Vieira, S., C. Stefanello, and J. Sorbara, *Formulating poultry diets based on their indigestible components*. **Poultry Science**, 2014. 93(9): p. 2411-2416.
153. Ravindran, V., *Feed enzymes: The science, practice, and metabolic realities*. **Journal of Applied Poultry Research**, 2013. 22(3): p. 628-636.
154. Bedford, M.a. and H. Schulze, *Exogenous enzymes for pigs and poultry*. **Nutrition Research Reviews**, 1998. 11(1): p. 91-114.
155. Bengtsson, S. and P. Åman, *Isolation and chemical characterization of water-soluble arabinoxylans in rye grain*. **Carbohydrate Polymers**, 1990. 12(3): p. 267-277.
156. Jürgens, H.-U., G. Jansen, and C.B. Wegener, *Characterisation of several rye cultivars with respect to arabinoxylans and extract viscosity*. **Journal of Agricultural Science**, 2012. 4(5): p. 1.
157. Mireles-Arriaga, A.I., et al., *Use of exogenous enzyme in animal feed*. **Life Science Journal**, 2015. 12(2): p. 23-32.
158. Courtin, C.M., et al., *Dietary inclusion of wheat bran arabinoxylooligosaccharides induces beneficial nutritional effects in chickens*. **Cereal Chemistry**, 2008. 85(5): p. 607-613.
159. Beauchemin, K., et al., *Use of exogenous fibrolytic enzymes to improve feed utilization by ruminants*. **Journal of Animal Science**, 2003. 81(14_suppl_2): p. E37-E47.
160. Krueger, N., et al., *The potential to increase digestibility of tropical grasses with a fungal, ferulic acid esterase enzyme preparation*. **Animal Feed Science and Technology**, 2008. 145(1-4): p. 95-108.
161. Addah, W., et al., *A third-generation esterase inoculant alters fermentation pattern and improves aerobic stability of barley silage and the efficiency of body weight gain of growing feedlot cattle1*. **Journal of Animal Science**, 2012. 90(5): p. 1541-1552.
162. Scalbert, A., et al., *Ether linkage between phenolic acids and lignin fractions from wheat straw*. **Phytochemistry**, 1985. 24(6): p. 1359-1362.
163. Muck, R.E., *Silage microbiology and its control through additives*. **Revista Brasileira de Zootecnia**, 2010. 39: p. 183-191.
164. Höök, M. and X. Tang, *Depletion of fossil fuels and anthropogenic climate change—A review*. **Energy Policy**, 2013. 52: p. 797-809.
165. Demirbas, A., *Biofuels sources, biofuel policy, biofuel economy and global biofuel projections*. **Energy Conversion and Management**, 2008. 49(8): p. 2106-2116.
166. Bhatia, R., et al., *Pilot-scale production of xylo-oligosaccharides and fermentable sugars from *Miscanthus* using steam explosion pretreatment*. **Bioresource Technology**, 2020. 296: p. 122285.
167. Somerville, C., *Biofuels*. **Current Biology**, 2007. 17(4): p. R115-R119.
168. de Vries, R.P., et al., *The current biotechnological status and potential of plant and algal biomass degrading/modifying enzymes from Ascomycete Fungi*. **Grand Challenges in Fungal Biotechnology**, 2020: p. 81-120.
169. Dodd, D. and I.K. Cann, *Enzymatic deconstruction of xylan for biofuel production*. **Gcb Bioenergy**, 2009. 1(1): p. 2-17.
170. Choudhary, J., et al., *Enhanced saccharification of steam-pretreated rice straw by commercial cellulases supplemented with xylanase*. **Journal of Bioprocessing & Biotechniques**, 2014. 4(7): p. 1.
171. Althuri, A. and S.V. Mohan, *Single pot bioprocessing for ethanol production from biogenic municipal solid waste*. **Bioresource Technology**, 2019. 283: p. 159-167.
172. Walia, A., et al., *Microbial xylanases and their industrial application in pulp and paper biobleaching: a review*. **3 Biotech**, 2017. 7(1): p. 11.
173. Buchert, J., et al., *Application of xylanases in the pulp and paper industry*. **Bioresource Technology**, 1994. 50(1): p. 65-72.
174. Kumar, V., et al., *Improved biobleaching of mixed hardwood pulp and process optimization using novel GA-ANN and GA-ANFIS hybrid statistical tools*. **Bioresource Technology**, 2019. 271: p. 274-282.

175. Zhao, L., et al., *Increasing efficiency of enzymatic hemicellulose removal from bamboo for production of high-grade dissolving pulp*. **Bioresource Technology**, 2017. 223: p. 40-46.
176. Willför, S., et al., *Polysaccharides in some industrially important softwood species*. **Wood Science and Technology**, 2005. 39(4): p. 245-257.
177. Record, E., et al., *Overproduction of the *Aspergillus niger* feruloyl esterase for pulp bleaching application*. **Applied Microbiology and Biotechnology**, 2003. 62(4): p. 349-355.
178. Kumar, N. and V. Pruthi, *Potential applications of ferulic acid from natural sources*. **Biotechnology Reports**, 2014. 4: p. 86-93.
179. Ou, S. and K.C. Kwok, *Ferulic acid: pharmaceutical functions, preparation and applications in foods*. **Journal of the Science of Food and Agriculture**, 2004. 84(11): p. 1261-1269.
180. Aachary, A.A. and S.G. Prapulla, *Xylooligosaccharides (XOS) as an emerging prebiotic: microbial synthesis, utilization, structural characterization, bioactive properties, and applications*. **Comprehensive Reviews in Food Science and Food Safety**, 2011. 10(1): p. 2-16.
181. Brooker, A.T., et al., *Catalytic laundry detergent composition comprising relatively low levels of water-soluble electrolyte*. **U.S. Patent Application**, No.2011/0005003.
182. Moid, M.M., et al. *Development of xylanase as detergent additive to improve laundry application*. **IOP Conference Series: Materials Science and Engineering**. 2021.
183. Whitaker, J.R., *Pectic substances, pectic enzymes and haze formation in fruit juices*. **Enzyme and Microbial Technology**, 1984. 6(8): p. 341-349.
184. Maillard, M.-N. and C. Berset, *Evolution of antioxidant activity during kilning: role of insoluble bound phenolic acids of barley and malt*. **Journal of Agricultural and Food Chemistry**, 1995. 43(7): p. 1789-1793.
185. Kumar, V., et al., *Global scenario of industrial enzyme market*. **Industrial Enzymes: Trends, Scope and Relevance**, 2014: p. 176-196.

章节二

通过生化分析未知分支的酯酶验证真菌
CE1两个亚家族生物功能



Chapter 2

Functional validation of two fungal subfamilies in Carbohydrate Esterase family 1 by biochemical characterization of esterases from uncharacterized branches

This chapter was published in *Frontiers in Bioengineering and Biotechnology*

Li, Xinxin, Kelli Griffin, Sandra Langeveld, Matthias Frommhagen, Emilie N. Underlin, Mirjam A. Kabel, Ronald P. de Vries, and Adiphol Dilokpimol

Volume 8, June 2020, 694

<https://doi.org/10.3389/fbioe.2020.00694>

Abstract

The fungal members of Carbohydrate Esterase family 1 (CE1) from the CAZy database include both acetyl xylan esterases (AXEs) and feruloyl esterases (FAEs). AXEs and FAEs are essential auxiliary enzymes to unlock the full potential of feedstock. They are being used in many biotechnology applications including food and feed, pulp and paper, and biomass valorization. AXEs catalyze the hydrolysis of acetyl group from xylan, while FAEs release ferulic and other hydroxycinnamic acids from xylan and pectin. Previously, we reported a phylogenetic analysis for the fungal members of CE1, establishing five subfamilies (CE1_SF1-SF5). Currently, the characterized AXEs are in the subfamily CE1_SF1, whereas CE1_SF2 contains mainly characterized FAEs. These two subfamilies are more related to each other than the other subfamilies and are predicted to have evolved from a common ancestor, but target substrates with a different molecular structure. In this study, four ascomycete enzymes from CE_SF1 and SF2 were heterologously produced in *Pichia pastoris* and characterized with respect to their biochemical properties and substrate preference towards different model and plant biomass substrates. The selected enzymes from CE1_SF1 only exhibited AXE activity, whereas the one from CE1_SF2 possessed dual FAE/AXE activity. This dual activity enzyme also showed broad substrate specificity towards model substrates for FAE activity and efficiently released both acetic acid and ferulic acid (~50%) from wheat arabinoxylan and wheat bran which was pre-treated with a commercial xylanase. These fungal AXEs and FAEs also showed promising biochemical properties, e.g., high stability over a wide pH range and retaining more than 80% of their residual activity at pH 6.0-9.0. These newly characterized fungal AXEs and FAEs from CE1 have high potential for biotechnological applications. In particular as an additional ingredient for enzyme cocktails to remove the ester-linked decorations which enables access for the backbone degrading enzymes. Among these novel enzymes, the dual FAE/AXE activity enzyme also supports the evolutionary relationship of CE1_SF1 and SF2.

Introduction

Over 5000 million tons of agro-food industrial side streams, such as wheat straw, rice straw, corn stover, potato peelings, and sugarcane bagasse, are produced from the agro-industry annually [1]. This plant biomass contains mainly polysaccharides, i.e., cellulose, hemicellulose and pectin, and the aromatic polymer lignin, which form a network of the plant cell wall and, therefore, provide protection against pathogens and pests [2, 3]. Xylan is a major component of hemicellulose from agro-food industrial side streams, accounting for 20-30% in secondary cell wall of dicots and up to 50% in commelinid monocots [4]. Xylan composes of a β -1,4-D-xylosyl backbone with different substituents, e.g., L-arabinose, D-galactose, D-(methyl)glucuronic acid, ferulic acid, and acetic acid. The acetyl groups are linked to the *O*-2 and/or *O*-3 position of the D-xylosyl units in commelinid monocots (cereals and grasses) [5-7], whereas feruloyl and to a lesser degree *p*-coumaryl groups are esterified mainly at the *O*-5 position on L-arabinosyl residues of xylan of commelinid monocots [8-11]. Feruloyl groups are also present in pectin, which esterified to the *O*-2 and/or *O*-5 positions of arabinan side chains and to the *O*-6 position of D-galactosyl residues in (arabino-)galactan of rhamnogalacturonan I [12-15]. Acetyl xylan esterases (AXEs, EC 3.1.1.72) catalyze the hydrolysis of ester linkages between acetyl groups and xylan, whereas feruloyl esterases (FAEs, EC 3.1.1.73) hydrolyze the ester linkages between hydroxycinnamic acids and plant cell-wall polysaccharides [16].

The Carbohydrate Active enZYme (CAZy) database classifies carbohydrate esterases into 17 families (CE1-CE17) [17]. The characterized fungal AXEs belong to CE1-CE6 and CE16 families, with the majority of the characterized enzymes in CE1. Part of the characterized fungal FAEs are also assigned to CE1, whereas the other FAE subfamilies are not classified in this database [18]. Recently, a phylogenetic classification of the fungal members of CE1 was established, which divided the fungal CE1 members into five subfamilies (CE1_SF1-SF5) [19]. The characterized AXEs were grouped in CE1_SF1 and the characterized FAEs were in CE1_SF2 and SF5, whereas none of the members from CE1_SF3 and SF4 has been characterized so far. Among these subfamilies, CE1_SF1 and SF2 are more related and originated from the same node, but target substrates with a different molecular structure. In this study, we aimed to explore the differences between CE1_SF1 and SF2 by selecting six ascomycete candidates from the uncharacterized branches of the CE1_SF1 and SF2 for heterologous production in *Pichia pastoris* and biochemical characterization using different model and plant biomass substrates.

Materials and methods

Materials

Methyl *p*-coumarate, methyl caffeate, methyl ferulate, methyl sinapate, ethyl *p*-coumarate, ethyl ferulate and chlorogenic acid were purchased from Apin Chemicals Limited (Oxon, United Kingdom). Insoluble wheat arabinoxylan (WAX, P-WAXYI, from wheat flour), endo- α -(1 \rightarrow 5)-arabinanase (E-EARAB, GH43 from *Aspergillus niger*), and endo- β -(1 \rightarrow 4)-galactanase (E-EGALN, GH53 from *A. niger*) were from Megazyme (Wicklow, Ireland). Corn oligosaccharides mix (COS) was provided by Wageningen University [7]. Wheat bran (WB) was from Wageningen Mill (Wageningen, The Netherlands). Corn xylooligosaccharides mix (CX) was from Carl Roth GmbH + Co. KG (Karlsruhe, Germany). Sugar beet pectin (SBP, Pectin Betapec RU301) was from Herbstreith & Fox KG (Neuenbürg, Germany). Xylanase (GH11 from *Thermomyces lanuginosus*), *p*-methyl umbelliferyl acetate (MUB-acetate) and other chemicals were purchased from Sigma-Aldrich (Merck KGaA, Darmstadt, Germany).

Bioinformatics

All amino acid sequences in this study were obtained from JGI MycoCosm (<https://genome.jgi.doe.gov/mycoCosm/home>) [20] and the CAZy database (<http://www.cazy.org/>) [7]. The secretory signal peptides were detected using the SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>) [21]. The amino acid sequences without predicted signal peptides were aligned using Multiple Alignment using Fast Fourier Transform (MAFFT, <https://mafft.cbrc.jp/alignment/server/>) [22] and visualized using Easy Sequencing in Postscript (<http://espript.ibcp.fr/ESPript/ESPript/>) [23]. The phylogenetic analysis was performed using maximum likelihood (ML), neighbor-joining (NJ) and minimal evolution (ME) implemented in the Molecular Evolutionary Genetic Analysis software version 7 (MEGA7, <https://www.megasoftware.net/>) [24] with 95% partial deletion and the Poisson correction distance of substitution rates. Statistical support for phylogenetic grouping was estimated by 500 bootstrap re-samplings. The final phylogenetic tree was shown using the ML tree with above 40% bootstrap next to the branches. *N*-glycosylation sites were predicted using NetNGlyc1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>) [25]. The theoretical molecular masses were calculated from the amino acid sequences without signal peptide using Sequence Manipulation Suite online tool (https://www.bioinformatics.org/sms2/protein_mw.html) [26].

Cloning and transformation to Pichia pastoris

The selected candidate genes (JGI accession numbers are provided in **Table 1**) without predicted signal peptide and introns were codon optimized and synthesized into pPicZ α A plasmid for production in *P. pastoris* (Genscript Biotech, Leiden, The Netherlands). The pPicZ α A containing synthetic genes were transformed into *Escherichia coli* DH5 α for propagation and sequencing. Then, the plasmids were extracted, linearized by *PmeI* (Promega, Madison, WI), and transformed into *P. pastoris* strain X-33 (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA) according to the manufacturer's recommendation [27]. The positive colonies were selected for the highest protein production based on colony Western Blot using anti Histidine-tag antibody conjugated with alkaline phosphatase (Thermo Fisher Scientific). The transformants were grown on a nitrocellulose membrane (0.45 μ m; Whatman, GE Healthcare Life Sciences, Buckinghamshire, United Kingdom) over minimal medium (MM, 1.34% yeast nitrogen base, 4×10^{-5} % biotin, 1% v/v methanol and 1.5% agar) for 2-4 days at 30°C. Then the membrane was washed with milliQ water, blocked with 2-5% skim milk in phosphate-buffered saline, and blotted with anti-Histidine-tag antibody. The signal was detected using 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium system.

Production of recombinant proteins

The selected *P. pastoris* transformants were grown according to Knoch et al. [28] in a buffered complex glycerol medium (BMGY, 1% yeast extract, 2% peptone, 0.1 M potassium phosphate buffer pH 6.0, 1% w/v glycerol, 1.34% yeast nitrogen base, 4×10^{-5} % biotin) overnight at 30°C, 250 rpm. Induction was done in buffered complex methanol medium (BMMY, BMGY without glycerol) at 22°C, 250 rpm with methanol supplemented (1% v/v) every 24 h for 96 h. Culture supernatants were harvested (4000 x g, 4°C, 20 min), filtered (0.22 μ m; Merck), aliquoted and stored at -20°C prior further analysis. The stability of the recombinant enzymes was assessed by their activity (see below) each time after thawing. The enzyme activity remained more than 95% of their original activity while performing the experiment (data not shown).

Biochemical properties of recombinant proteins

Molecular masses of the recombinant proteins were estimated by SDS-PAGE (12% w/v, sodium dodecyl sulfate-polyacrylamide gel) using Mini-PROTEAN® Tetra Cell (Bio-Rad, Hercules, CA). Deglycosylation was performed by incubating 20 μ L of *P. pastoris* culture supernatant with endoglycosidase H (New England Biolabs, Ipswich, MA) as recommended by the manufacturer.

Protein concentration was assessed from SDS-PAGE gel by densitometric method using ImageJ program [29] and 0.5-2.0 μg Bovine Serum Albumin (Pierce, Thermo Scientific, Carlsbad, CA) as a standard.

Enzyme Activity Assays

p-Methyl umbelliferyl substrate assay

Activity towards *p*-methylumbelliferyl acetate (MUB-acetate) was performed in 100 μL reaction mixtures containing 10 μM MUB- acetate (dissolved in acetone), 100 mM phosphate buffer (pH 6.0) and 10 μL culture supernatant. The reaction was performed at 30°C. The release of umbelliferone group was spectrophotometrically quantified by following the excitation/emission at 340/520 nm up to 30 min with a 2 min interval. The activity of the enzymes was determined by quantification of the umbelliferone group using a standard curve (0.5-100 μM). One unit (U) of AXE activity was defined as the amount of enzyme which released 1 μmol of umbelliferone group from MUB-acetate per min under the assay condition. Culture filtrate from *P. pastoris* harboring pPicZaA plasmid without insertion was used as negative control. All assays were performed in triplicate.

Hydroxycinnamate substrate assay

Activity of the CE1_SF1 and SF2 candidates towards hydroxycinnamate substrates (methyl *p*-coumarate, methyl caffeate, methyl ferulate, methyl sinapate, ethyl *p*-coumarate, ethyl ferulate and chlorogenic acid) was performed in 250 μL reaction mixtures containing 0.12 mM substrate (dissolved in dimethylformamide), 80 mM phosphate buffer (pH 6.0) and 50 μL culture supernatant. The decrease of substrate concentration was spectrophotometrically quantified by following the absorbance at 340 nm up to 30 min with a 2 min interval. The activity of the enzymes was determined by quantification of the substrate concentration using standard curves in a range between 0.005-0.25 mM. One unit (U) of FAE activity was defined as the amount of enzyme which decreased 1 μmol of substrates per min under the assay condition. Culture filtrate from *P. pastoris* harboring pPicZaA plasmid without insertion was used as negative control. All assays were performed in triplicate.

pH and temperature profiles

MUB-acetate and methyl ferulate were used as substrates for AXE and FAE activity, respectively, based on the above assay. pH profiles of enzyme activity were determined at 30°C in 100 mM Britton-Robinson buffer [30] (pH 2.0-9.0). The temperature profiles of enzyme activity were measured in 100 mM phosphate buffer pH 6.0 at 22-80°C. The pH and thermal stability of the enzymes were determined by measuring the residual enzyme activities after 2 h or 16 h incubation at 30°C in 100 mM Britton-Robinson buffer pH 2.0-10.0, or after 2 h incubation at 22°C to 80°C in phosphate buffer pH 6.0 [31].

Hydrolytic activity towards (poly- /oligo-)saccharides

The activity towards plant biomass was determined using insoluble wheat arabinoxylan (WAX) and corn oligosaccharides mix (COS) for AXE activity, and WAX, wheat bran (WB), corn xylooligosaccharides mix (CX) and sugar beet pectin (SBP) for FAE activity. For pre-treatment, 1% of WAX, COS, WB and CX was incubated with 0.1 mg xylanase, or 0.1 mg of arabinanase and galactanase for SBP, in a 50 mM sodium acetate buffer pH 4.5 and 0.02% sodium azide at 100 rpm and 30°C for 72 h. The reaction was stopped by heat inactivation at 95°C for 10 min [27]. For hydrolytic activity assay, the reaction containing 500 μL 1% (pre-treated) substrate and 100 μL culture supernatant (containing 1 μg enzyme) was incubated at 30°C, 24 h, 100 rpm. Except for COS reactions that were stopped by the addition of 50 μL 2 M HCl, the others were inactivated by heating at 95°C, 10 min and centrifuged at 14,000 rpm, 4°C for 15 min. The supernatants were used for HPLC analysis.

Acetic acid content analysis

The acetic acid release was measured by using HPLC (Dionex ICS-3000 chromatography system; Thermo Scientific, Sunnyvale, CA) equipped with an Aminex HPX 87H column with a guard-column (300 x 7.8 mm; Bio-Rad) and a refractive index detector (Bio-Rad). An isocratic elution of 5.0 mM sulfuric acid, with a flow rate of 0.6 mL/min at 40°C was used according to Zeppa et al. [32]. The release of acetic acid was quantified using standards in a range between 0.01-2.0 mg/mL.

Hydroxycinnamic acid content analysis

The reaction supernatants were mixed with 100% acetonitrile (1:3, v/v), then incubated on ice for 10 min and centrifuged for 15 min at 4°C to remove precipitants prior to the analysis. The ferulic release was monitored by using HPLC (Dionex ICS-5000+ chromatography system; Thermo Scientific, Sunnyvale, CA) equipped with an Acclaim Mixed-Mode WAX-1 LC Column (3 x 150 mm; Thermo Scientific) and a UV detector (310 nm; Thermo Scientific). The chromatographic separation was carried out according to Dilokpimol et al. [27] using an isocratic elution of 25 mM potassium phosphate buffer, 0.8 mM pyrophosphate, pH 6.0 in 50% (v/v) acetonitrile with a flow rate of 0.25 mL/min at 30°C. Ferulic and *p*-coumaric acids (0.25-50 µM) were used as standards for

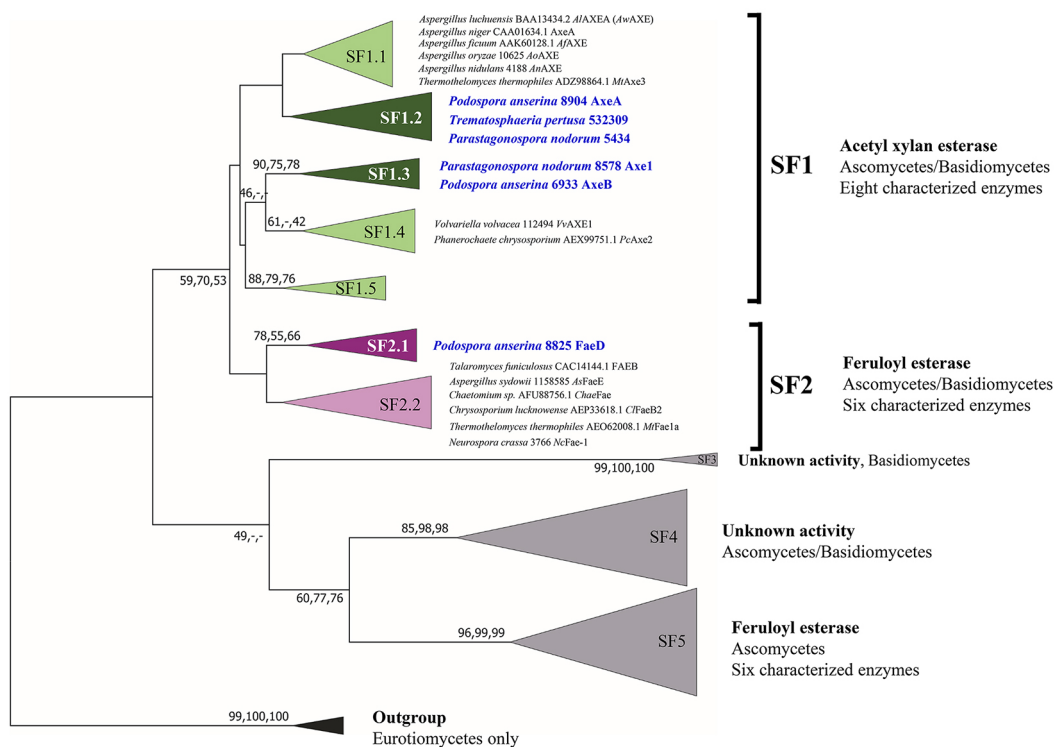


Figure 1 Reconstructed phylogenetic tree of fungal CE1 members (modified from Mäkelä et al. [19]). The phylogenetic analysis was performed by maximum likelihood (ML) implemented in MEGA7 [24] with 95% partial deletion of gaps and the Poisson correction distance of substitution rates. The main branches/subfamilies were collapsed. Statistical support for phylogenetic grouping was estimated by 500 bootstrap re-samplings, only the bootstrap above 40% were shown on the branches, also of a neighbor-joining (NJ) and minimal evolution (ME) tree using the same dataset (order: ML, NJ, ME). Eight FAEs from subfamily 7 [18] were used as an outgroup. Names in blue and bold indicates the selected candidates for this study. SF, subfamily. The full phylogenetic tree can be found in **Suppl. Fig. 1**.

identification and quantification.

Results

Production of four new fungal CE1 enzymes

Previous phylogenetic tree for fungal CE1 was purposed which classified the members of CE1 into five subfamilies [19]. Using a genome-mining strategy by BLAST searching against 18 genome sequenced fungal species [33], we added 87 additional CE1 sequences to the previous CE1 tree [19], and reconstructed the phylogenetic tree (**Figure 1, Suppl. Fig. 1, Suppl. Table 1**). The reconstructed CE1 tree is agreed with the previous one. CE1_SF1 could be further separated into five branches (CE1_SF1.1-SF1.5), while CE1_SF2 was divided into two branches (CE1_SF2.1-SF2.2). Fungal CE1_SF1 contains nine biochemically characterized AXEs from CE1_SF1.1 and SF1.4, while fungal CE1_SF2 contains six biochemically characterized FAEs that were all from CE1_SF2.2 (**Figure 1, Table 1**). It is possible that the uncharacterized branches contain enzymes with different activity. To systematically evaluate the subfamilies and compare the activities and biochemical properties, six fungal candidates from the uncharacterized branches of these two subfamilies (three from CE1_SF1.2, two from CE1_SF1.3 and one from CE1_SF2.1) were selected for recombinant protein production in *P. pastoris*.

Of these six candidates, only four candidate enzymes, i.e., *Podospora anserina* AxeA, AxeB, *Parastagonospora nodorum* Axe1, and *Po. anserina* FaeD, were detected by SDS-PAGE (**Suppl. Fig. 2**). The apparent masses of *Pa. nodorum* Axe1, *Po. anserina* AxeB, and *Po. anserina* FaeD were 40, 100, and 45 kDa, respectively, whereas *Po. anserina* AxeA showed smears. After deglycosylation using endoglycosidase H, the molecular masses of *Po. anserina* AxeA, *Pa. nodorum* Axe1, and *Po. anserina* FaeD decreased to 35, 35, and 35 kDa, respectively, which corresponded with their calculated masses based on the amino acid sequence. In contrast, the deglycosylation of *Po. anserina* AxeB still resulted in a molecular mass of about 70 kDa, which was still higher than the calculated molecular mass. It is possible that the glycosylation sites of *Po. anserina* AxeB were inaccessible for endoglycosidase H or that other post-translational modifications were present.

All four new enzymes have high alkaline tolerance

Po. anserina AxeA, AxeB, *Pa. nodorum* Axe1 were active toward MUB-substrates, while *Po. anserina* FaeD showed a broad substrate range. It hydrolyzed all tested substrates and showed the highest activity toward methyl *p*-coumarate (141.5 ± 3.5 U/mg) (**Table 2**).

Based on MUB-acetate, *Po. anserina* AxeA, AxeB, *Pa. nodorum* Axe1 and *Po. anserina* FaeD showed the highest activity at pH 7.0, 6.0, 7.0, and 7.0, respectively, when incubated at 30°C (**Figure 2A**). They showed the highest activity at 40, 40, 50, and 30°C, respectively, when incubated at pH 6.0 (**Figure 2B**). The enzymes showed stability over a wide pH range, retaining more than 80% residual activity after incubation at pH 6.0-9.0 for 2 h (**Figure 2C**). In addition, *Pa. nodorum* Axe1 and *Po. anserina* FaeD retained more than 85% residual activity between pH 3.0 and 10.0. The stability of *Po. anserina* AxeA and *Pa. nodorum* Axe1 was higher, since both enzymes retained 55 and 90% of their residual activity at pH 2, compared to *Po. anserina* AxeB and *Po. anserina* FaeD, which did not show any residual activity after incubation for 16 h (**Figure 2E**). Most enzymes retained more than 80% residual activity after incubation at 40°C for 2 h except for *Po. anserina* AxeB that retained only 55% residual activity (**Figure 2D**).

Furthermore, *Po. anserina* FaeD showed the highest activity towards methyl ferulate at pH 7.0 and 50°C (**Figure 3A, 3B**). It retained more than 80% residual activity after incubation at pH 4.0-10.0 for 2 h (**Figure 3C**). *Po. anserina* FaeD also retained more than 70% residual activity after incubation at 80°C for 2 h (**Figure 3D**).

Table 1 Molecular mass and properties of characterized CE1_SF1 and SF2 and the selected enzymes in this study.

SF	Fungal species	GenBank or JGI accession number	Enzyme name	Calculated molecular mass (kDa)	Apparent molecular mass (kDa) ^{a,b}	pH optimum ^a	pH stability ^a	Temp. optimum (°C) ^a	Temp. stability (°C) ^a	References
1.1	<i>Aspergillus luchuensis</i> ^c	BAA13434.2	<i>A</i> /AXEA (<i>Aw</i> AXE)	33	30	7.0	6.0-9.0	N.A.	40	[34, 35]
1.1	<i>Aspergillus niger</i>	CAA01634.1	AxeA	33	N.A.	N.A.	N.A.	N.A.	N.A.	[36]
1.1	<i>Aspergillus ficuum</i>	AAK60128.1	<i>A</i> /AXE	33	34	7.0	N.A.	37	50	[37]
1.1	<i>Aspergillus oryzae</i>	jgi Aspor1 10625	<i>Ao</i> AXE	34	30	6.0	6.0-7.0	45	40-50	[38]
1.1	<i>Aspergillus nidulans</i>	jgi Aspnid1 4188	<i>An</i> AXE	34	N.A.	N.A.	N.A.	N.A.	N.A.	[39]
1.1	<i>Thermothelomyces thermophilus</i> ^d	ADZ98864.1	<i>M</i> AXe3	35	N.A.	7.0	N.A.	35-45	N.A.	[40]
1.1	<i>Rasamsonia emersonii</i>	ADX07526.1	<i>T</i> eCE1	30	33	N.A.	N.A.	N.A.	N.A.	[41]
1.2	<i>Podospora anserina S mat+</i>	jgi Podan2 8904	AxeA	32	35	7.0	6.0-9.0^e	40	40^f	This study
1.3	<i>Parastagonospora nodorum S/N1</i> ^g	jgi Stamo2 8578	Axe1	35	35	7.0	6.0-9.0^e	30	40^f	This study
1.3	<i>Podospora anserina S mat+</i>	jgi Podan2 6933	AxeB	45	70	6.0	6.0-9.0^e	40	30^f	This study
1.4	<i>Vohvarella vohvocea</i>	jgi Vohvo 112494	<i>Vv</i> AXE1	38	45	8.0	N.A.	60.0	40-50	[42]
1.4	<i>Phanerochaete chrysosporium</i>	AEX99751.1	<i>Pc</i> Axe2	40	55	7.0	N.A.	30-35	N.A.	[43]
2.1	<i>Podospora anserina S mat+</i>	jgi Podan2 8825	FaeD	33	35	7.0	3.0-10.0^e	50	40^f	This study
2.2	<i>Penicillium funiculosum</i>	CAC14144.1	FAEB	39	35	N.A.	N.A.	N.A.	N.A.	[44]
2.2	<i>Aspergillus sydowii</i>	jgi Aspsy1 1158585	<i>As</i> FaeE	33	32	N.A.	N.A.	N.A.	N.A.	[33]
2.2	<i>Neurospora crassa</i>	jgi Neuer2 3766	<i>Nc</i> Fae-1	32	29	6.0	6.0-7.5	55	N.A.	[45]
2.2	<i>Chaetomium sp.</i>	AFU88756.1	<i>Chae</i> Fae	33	30	7.5	4.0-10.0	60	55	[46]
2.2	<i>Thermothelomyces thermophilus</i> ^d	AEO62008.1	<i>M</i> Fae1a	33	30	7.0	7.0-10.0	50	55	[47]
2.2	<i>Chrysosporium lucknowense</i>	AEP33618.1	<i>C</i> FaeB2	33	33	7.0	N.A.	45	N.A.	[48]

a, Not available.

b, Apparent mass after deglycosylation by Endoglycosidase H.

c, Formerly *Aspergillus awamori*.d, Formerly *Myceliophthora thermophila*.e, Formerly *Stagonospora/Septoria nodorum*.

f, Retaining more than 80% residual activity after incubating at 30°C or at pH 6.0 for pH or temperature stability, respectively.

Enzymes in bold are used in this study.

Table 2 Specific activity of positive candidates toward model substrates.

Fungal species	Enzyme	Concentration (mg/L)	Specific activity (U/mg) ^a								
			MUB-acetate	Methyl <i>p</i> -coumarate	Methyl caffeate	Methyl ferulate	Methyl sinnapate	Ethyl <i>p</i> -coumarate	Ethyl ferulate	Chlorogenic acid	
<i>Podospora anserina</i>	AxeA	12	30.4 ± 0.4	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Parastagonospora nodorum</i>	AxeB	21	33.4 ± 0.3	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Podospora anserina</i>	Axe1	55	22.1 ± 0.0	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
<i>Podospora anserina</i>	FaeD	73	3.1 ± 0.1	141.5 ± 3.5	53.0 ± 0.3	49.8 ± 0.3	50.0 ± 0.1	48.1 ± 2.1	39.2 ± 0.6	22.8 ± 1.4	

a, Not active.

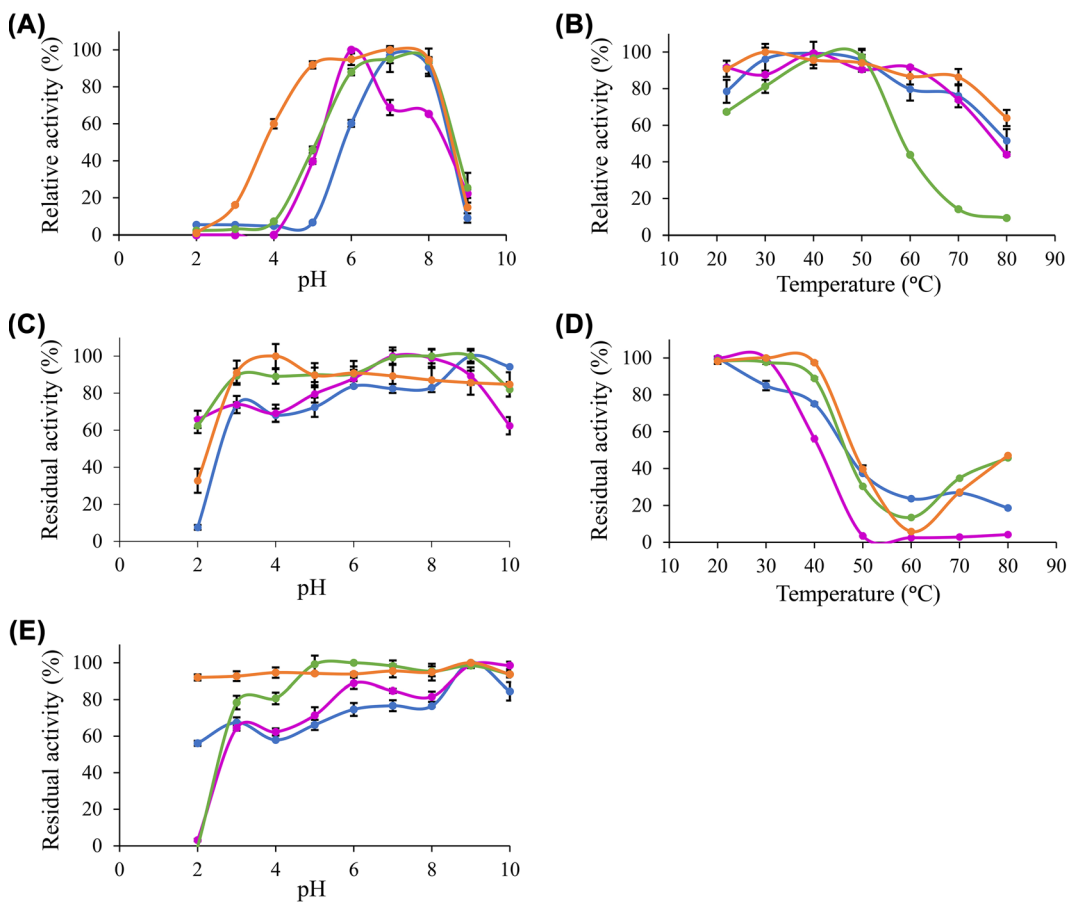


Figure 2 pH and temperature profiles of the selected CE1 enzymes toward *p*-methylumbelliferyl acetate. (A) pH optimum; (B) Temperature optimum; (C) pH stability for 2 h incubation at 30°C; (D) Temperature stability for 2 h incubation at pH 6.0; (E) pH stability for 16 h incubation at 30°C. ● *Po. anserina* AxeA; ● *Pa. nodorum* Axe1; ● *Po. anserina* AxeB; ● *Po. anserina* FaeD. Each experiment was performed in triplicate. Standard deviations are shown as error bars.

Po. anserina FaeD is a dual FAE/AXE activity enzyme

WAX and COS were used as substrates to determine the ability of the enzymes to release acetic acid from plant biomass (**Figure 4**). WAX and COS contain about 1 and 3.2 mg acetic acid per one g polysaccharide, respectively. *Po. anserina* AxeA, *Pa. nodorum* Axe1 and *Po. anserina* FaeD released acetic acid from both WAX and COS, whereas *Po. anserina* AxeB only released acetic acid from WAX. The highest acetic acid release was detected for *Po. anserina* AxeA and *Pa. nodorum* Axe1 from WAX (78%) and COS (85%), respectively. A non-CE1 FAE from *Penicillium subrubescens* FaeA [18] was used for comparing the degrading ability of CE1 enzymes. *P. subrubescens* FaeA showed no release of acetic acid from either WAX or COS.

WAX, WB, CX and SBP were used as substrates to verify the ability of the enzymes to release ferulic acid from plant biomass (**Figure 5**), as these substrates contain about 3, 1, 40 and 1.9 mg total ferulic acid per one g polysaccharide, respectively [27]. *Po. anserina* FaeD released ferulic acid from all tested substrates, whereas none of the others showed any FAE activity. *Po. anserina* FaeD released the highest amount (64%) of ferulic acid from CX which was pre-treated with xylanase and no ferulic acid was detected when CX was not pre-treated with xylanase. *Po. anserina* FaeD

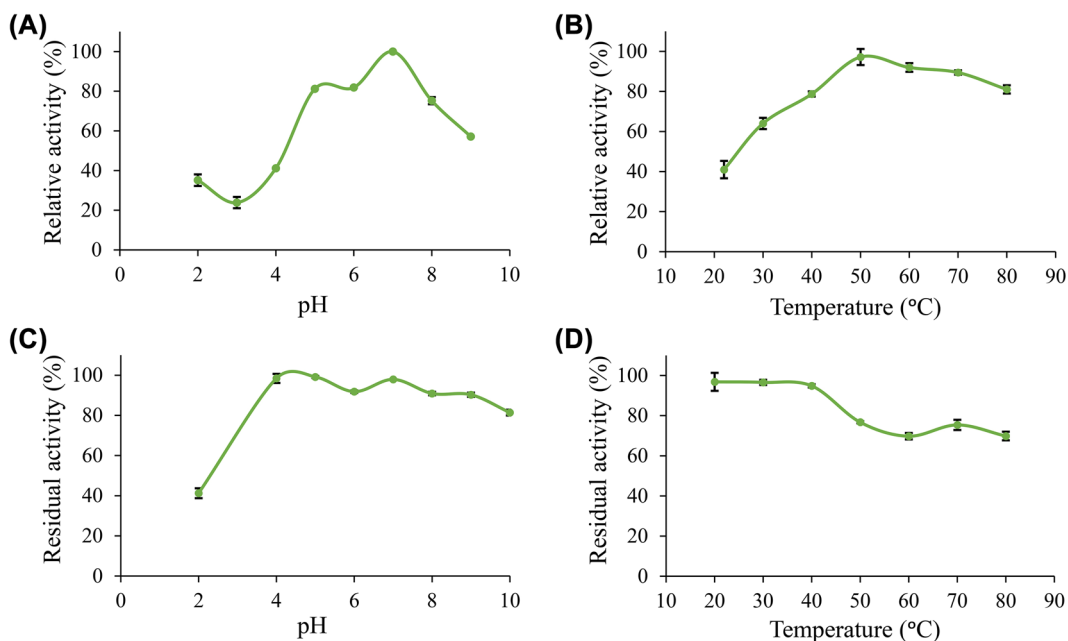


Figure 3 pH and temperature profiles of *Po. anserina* FaeD toward methyl ferulate. (A) pH optimum; (B) Temperature optimum; (C) pH stability for 2 h incubation at 30°C; (D) Temperature stability for 2 h incubation at pH 6.0. Each experiment was performed in triplicate. Standard deviations are shown as error bars.

also released a higher amount of ferulic acid from pre-treated WAX and WB with xylanase (54% and 48%, respectively) than those without xylanase pre-treatment (2% and 8%). *P. subrubescens* FaeA released a higher amount of ferulic acid from WAX, WB and CX than *Po. anserina* FaeD, and released all ferulic acid from pre-treated WAX and WB. The incubation of SBP with arabinanase and galactanase did not improve the release of ferulic acid by *Po. anserina* FaeD and *P. subrubescens* FaeA, and both enzymes only released less than 10% ferulic acid from SBP.

Discussion

The CE1 is a diverse family and the fungal CE1 members can be classified into five subfamilies (CE1_SF1-SF5) (**Figure 1 and Suppl. Fig. 1**). Multiple sequence alignment of members from CE1_SF1 and SF2 showed that these enzymes share an amino acid sequence similarity above 70%, which indicated that both CE1_SF1 and SF2 are more related to each other than to the other subfamilies. The same catalytic triad (Ser/Asp/His), the signature motif (G-X-S-X-G) and the shared common node suggested that these two subfamilies are likely to evolve from the same ancestor [19, 34, 35]. However, both subfamilies target substrates with a different molecular structure. So far, only a limited number of fungal CE1 enzymes have been characterized (**Table 1**), which hinders to explain the different activity and evolution relationship between CE1_SF1 and SF2.

In this study, we selected six ascomycete candidates from the uncharacterized branches in CE1_SF1 and SF2 to further explore the differences between CE1_SF1 and SF2. Four of these (*Po. anserina* AxeA, *Po. anserina* AxeB, *Pa. nodorum* Axe1 and *Po. anserina* FaeD) could be successfully produced in *P. pastoris* and exhibited different properties. Both *Pa. nodorum* and *Po. anserina* belong to Phylum Ascomycota, Subphylum Pezizomycotina, which the first is classified in the Class Dothideomycetes, the latter is part of the Class Sordariomycetes. Based on the AXE activity assay, *Po. anserina* AxeA, *Po. anserina* AxeB and *Pa. nodorum* Axe1 from CE1_SF1 showed the

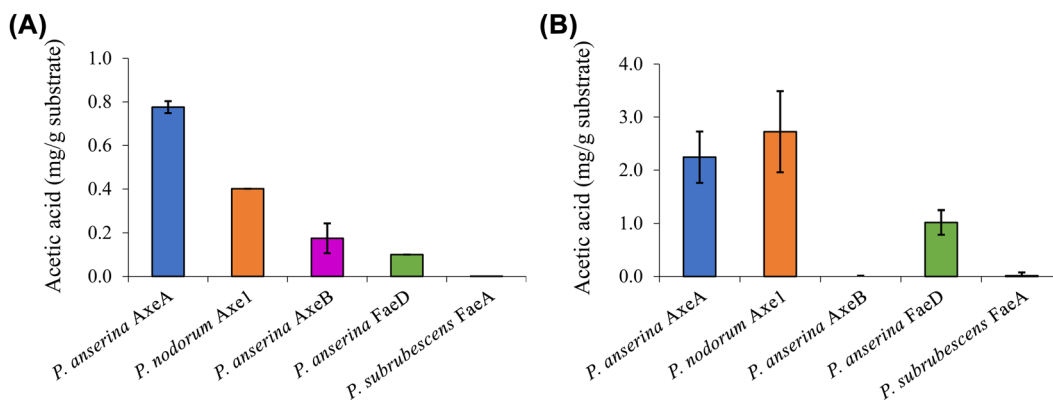


Figure 4 Acetic acid release from wheat arabinoxylan and corn oligosaccharides the selected CE1 enzymes or *P. subrubescens* FaeA. (A) Wheat arabinoxylan; (B) corn oligosaccharides. Each experiment was performed in triplicate. Standard deviations are shown as error bars.

highest activity toward MUB-acetate, while *Po. anserina* FaeD from CE1_SF2 showed around 10-fold less activity on this substrate (**Table 2**). Even though all four enzymes showed activity toward AXE substrates, only *Po. anserina* FaeD also de-esterified the FAE substrates with broad substrate preference. Most characterized FAEs from this subfamily showed broad substrate specificity including FaeB, *AsFae*, *MtFae1a* and *C/FaeB2*, supporting that this is one of the unique features of the FAEs in CE1 [18, 33, 36-38].

All enzymes could release acetic acid from both WAX and COS, except for *Po. anserina* AxeB for which the release of acetic acid was only detected from WAX (**Figure 4**). This could result from the different position of the acetylation in the substrates. Acetylation in commelinid monocots including wheat generally occurs at the *O*-2 and/or *O*-3 position of the D-xylosyl residues of the xylan backbone [5, 6]. It has been shown earlier that AXEs from CE1 regioselectively cleave the substituents in the *O*-2 and *O*-3 position, and deacetylate the *O*-2 position faster than the *O*-3 position [39]. However, COS contains more than half of the acetyl group attached to the *O*-2 position of the D-xylosyl backbone, which the same D-xylose is also substituted with a monomeric α -L-arabinosyl residue at the *O*-3 position [7]. Because *Po. anserina* AxeA and *Pa. nodorum* Axe1 released over 75% of total acetic acid, our results also indicated the potential ability of these enzymes to attack the dense acetyl substitution on the xylan backbone in COS.

In contrast, only *Po. anserina* FaeD could release ferulic acid from the feruloylated substrates (**Figure 5**). As mentioned before, feruloylation in commelinid monocots mainly occurs at the *O*-5 position on L-arabinosyl residues of xylan [8-11], while in pectin it mainly occurs at the *O*-2 and/or *O*-5 positions of arabinan side chains and at the *O*-6 position of D-galactosyl residues in (arabino-) galactan of rhamnogalacturonan I [12-15]. *Po. anserina* FaeD was shown to release more ferulic acid from commercial xylanase-treated xylan substrates (WAX, WB and CX) than non-treated ones, from which it released only small amounts of ferulic acid. In contrast, pre-treatment of SBP with arabinanase and galactanase did not improve the release of ferulic acid by *Po. anserina* FaeD and less than 10% of ferulic acid was released from intact and pre-treated pectin. Analysis of ferulic acid release from xylan and SBP substrates indicated that pre-treatment had a larger effect on *Po. anserina* FaeD activity on xylan substrates than SBP. This difference indicated that *Po. anserina* FaeD is mainly active on *O*-5 feruloylated xylooligosaccharides and much less on other substitutions. *Po. anserina* FaeD from CE1_SF2 showed a high preference for xylan substrates, which resembles FAEB from *Penicillium funiculosum* [36] and FaeC from *Aspergillus niger* [27]. *P. subrubescens* FaeA, a non-CE1 FAE control, did not release acetic acid from WAX or COS, but could efficiently release ferulic acid from WAX, WB and CX. Earlier, an FAE from *Aspergillus oryzae* (*AoFae*) was shown to possess dual activity [40]. *AoFae* also belongs to CE1 based on sequence homology of the

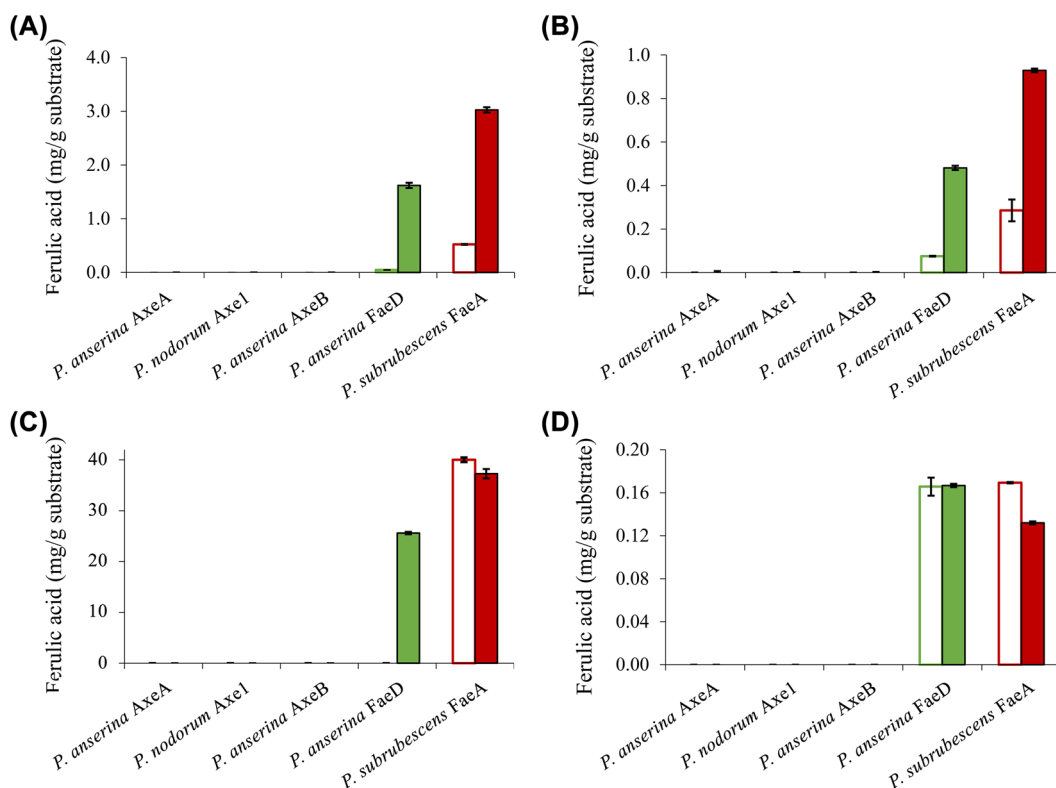


Figure 5 Ferulic acid release from plant biomass using the selected CE1 enzymes or *P. subrubescens* FaeA. (A) Wheat arabinoxylan; (B) wheat bran; (C) corn xylooligosaccharides; (D) sugar beet pectin. Empty bars indicate reactions with non-treated plant biomass, filled bars indicate reactions with pre-treated biomass (with xylanase for wheat arabinoxylan, wheat bran, corn xylooligosaccharides or arabianase and galactanase for sugar beet pectin) prior to incubation with CE1 enzymes or *P. subrubescens* FaeA. Each experiment was performed in triplicate. Standard deviations are shown as error bars.

N-terminal amino acid sequence [41]. However, based on the hydrolysis pattern toward different monoacetylated and monoferuloylated *p*-nitrophenyl glycosides, AoFae was later suggested to be a non-specific acetyl esterase [42]. In the same study, the authors also showed that some FAEs could also release acetyl residue from the same substrates with specific positional specificity, but with at least two orders of magnitude lower [42]. This aspect should be validated further for these new enzymes to monitor the de-esterified positions.

Based on the multiple sequence alignment (Suppl. Fig. 3) between CE1_SF1 and SF2, Trp160 was highly conserved among CE1_SF1 members, whereas the corresponding amino acids in CE1_SF2 can be an Ala, Ser, Pro, Gln or Thr (Suppl. Table 1). Recently, a crystal structure of an *Aspergillus luchuensis* (formerly *A. awamori*) AlAXEA (PDB code 5X6S) showed Trp160 controls the substrate specificity of AXE in CE1 [35]. When replacing Trp160 to Ala, Ser, or Pro, the mutants showed significant FAE activity. Trp160 in AlAXEA is corresponding to an Ala in *Po. anserina* FaeD supporting that the expanded substrate specificity of an AXE to FAE and the dual activity is potentially influenced by this amino acid.

To investigate the pH and temperature profile of these recombinant enzymes, we used MUB-acetate as a substrate for AXE activity. All candidates showed optimum pH in a neutral range (pH 6.0 to 7.0) and an optimum temperature between 30 and 50°C, which are similar to most other characterized CE1_AXE enzymes [34, 43-46] (Table 1). Surprisingly, they all showed excellent pH stability over

a broad pH range, retaining more than 80% residual activity after an incubation at pH 6.0-9.0 for 2 h, which showed a similar pH stability range to *Aspergillus luchuensis* ALAXEA [34, 35]. It should be noted that at pH more than 7.0, the ester-linkages are alkali labile and tend to degrade easily. Hence, to determine the pH stability of AXEs, we used culture filtrate from *P. pastoris* harboring pPicZaA plasmid without insertion as negative control, from which the residual activity was deduced. Using methyl ferulate as a substrate for FAE activity, *Po. anserina* FaeD also showed optimum pH at 7.0 and optimum temperature at 50°C, these properties are within the range of other reported fungal FAEs, mainly with activity at pH 4-8 and temperature 30-65°C [37, 38, 47, 48] (**Table 1**). It also showed a high pH stability over a broad pH range from pH 3.0 to 10.0, which is more superior than most reported CE1 enzymes (**Table 1**). The CAZymes from *Pa. nodorum* (synonyms: *Stagonospora/Septoria/Phaeosphaeria nodorum*) have not been characterized in much detail, mainly because it is a major pathogen of wheat and related cereals [49]. Several CAZymes from *Po. anserina* were previously reported, e.g., *PaMan5A* and *PaMan26A* mannanases, *PaXyn11A* xylanase, and *PaAbf51A* and *PaAbf62A* arabinofuranosidases were active in a range of pH 3-7 and 5-75°C [50], while *PaCel6A-C* were active in a range of pH 4-9 and 25-65°C [51]. The biochemical properties of different enzymes from the same strain can be vastly different also depending on the function of the enzyme, which most of the enzymes in this study are quite stable at high pH. The alkaline tolerance of the newly characterized enzymes is of interested for many bio-industrial applications particularly for the alkali pre-treated plant biomass.

Conclusion

CE1_SF1 and SF2 are related, even though the characterized enzymes from the first possess AXE activity and the ones from latter possess FAE activity. So far no additional activity was detected in these subfamilies except for the dual activity. A novel dual FAE/AXE activity enzyme was identified from CE1_SF2, which showed promising industrial applications because of its broad substrate specificity. To further understand the functional features and physiological role of individual enzymes, the positional specificity of these new enzymes should be further investigated. The phylogenetic analysis and multiple sequence alignment supported the evolutionary relationship between CE1_SF1 and SF2, and identified possible amino acids that control the AXE or FAE activity between these two subfamilies. Moreover, these novel fungal AXE and FAE enzymes showed a great alkaline tolerance and can selectively release acetic acid and ferulic acid from different plant-based biomass making them attractive for various biotechnological applications.

Author contributions

XL, KG, AD conducted the main experiments, analyzed the data, and draft the manuscript. MF, ENU, MAK contributed to HPLC analysis for acetic acid. XL, KG, SL contributed to phylogenetic analysis. AD, RPDV contributed to data interpretation and commented on the manuscript. RPDV conceived and supervised the overall project. All authors commented on the manuscript.

Funding

This project was partly supported by China Scholarship Council (Beijing, China), grant number: 201803250066.

References

1. Chandra, R., H. Takeuchi, and T. Hasegawa, *Methane production from lignocellulosic agricultural crop wastes: A review in context to second generation of biofuel production*. **Renewable and Sustainable En-**

- ergy Reviews*, 2012. 16(3): p. 1462-1476.
2. Ithal, N., et al., *Developmental transcript profiling of cyst nematode feeding cells in soybean roots. Molecular Plant-Microbe Interactions*, 2007. 20(5): p. 510-525.
 3. Liu, Q., L. Luo, and L. Zheng, *Lignins: biosynthesis and biological functions in plants. International Journal of Molecular Sciences*, 2018. 19(2): p. 335.
 4. Scheller, H.V. and P. Ulvskov, *Hemicelluloses. Annual Review of Plant Biology*, 2010. 61: p. 263-289.
 5. Puls, J. *Chemistry and biochemistry of hemicelluloses: Relationship between hemicellulose structure and enzymes required for hydrolysis. Macromolecular Symposium*, 1997. 120: p. 183-196.
 6. Appeldoorn, M.M., et al., *Characterization of oligomeric xylan structures from corn fiber resistant to pre-treatment and simultaneous saccharification and fermentation. Journal of Agricultural and Food Chemistry*, 2010. 58(21): p. 11294-11301.
 7. Appeldoorn, M.M., et al., *Enzyme resistant feruloylated xylooligomer analogues from thermochemically treated corn fiber contain large side chains, ethyl glycosides and novel sites of acetylation. Carbohydrate Research*, 2013. 381: p. 33-42.
 8. Smith, M.M. and R.D. Hartley, *Occurrence and nature of ferulic acid substitution of cell-wall polysaccharides in graminaceous plants. Carbohydrate Research*, 1983. 118: p. 65-80.
 9. Saulnier, L., J. Vigouroux, and J.-F. Thibault, *Isolation and partial characterization of feruloylated oligosaccharides from maize bran. Carbohydrate Research*, 1995. 272(2): p. 241-253.
 10. Wende, G. and S.C. Fry, *O-feruloylated, O-acetylated oligosaccharides as side-chains of grass xylans. Phytochemistry*, 1997. 44(6): p. 1011-1018.
 11. Schendel, R.R., et al., *Feruloylated Wheat Bran Arabinoxylans: Isolation and Characterization of Acetylated and O-2-Monosubstituted Structures. Cereal Chemistry*, 2016. 93(5): p. 493-501.
 12. Carpita, N.C., *Structure and biogenesis of the cell walls of grasses. Annual Review of Plant Biology*, 1996. 47(1): p. 445-476.
 13. Levigne, S.V., et al., *Isolation from sugar beet cell walls of arabinan oligosaccharides esterified by two ferulic acid monomers. Plant Physiology*, 2004. 134(3): p. 1173-1180.
 14. Harris, P., *Plant diversity and evolution: genotypic and phenotypic variation in higher plants. Diversity in plant cell walls*, 2005: p. 201-227.
 15. Harris, P.J. and J.A. Trethewey, *The distribution of ester-linked ferulic acid in the cell walls of angiosperms. Phytochemistry Reviews*, 2010. 9(1): p. 19-33.
 16. Christov, L.P. and B.A. Prior, *Esterases of xylan-degrading microorganisms: production, properties, and significance. Enzyme and Microbial Technology*, 1993. 15(6): p. 460-475.
 17. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research*, 2014. 42(D1): p. D490-D495.
 18. Dilokpimol, A., et al., *Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. Biotechnology for Biofuels*, 2016. 9(1): p. 1-18.
 19. Mäkelä, M.R., et al., *Characterization of a feruloyl esterase from Aspergillus terreus facilitates the division of fungal enzymes from Carbohydrate Esterase family 1 of the carbohydrate-active enzymes (CAZy) database. Microbial Biotechnology*, 2018. 11(5): p. 869-880.
 20. Grigoriev, I.V., et al., *MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research*, 2014. 42(D1): p. D699-D704.
 21. Petersen, T.N., et al., *SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods*, 2011. 8(10): p. 785-786.
 22. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution*, 2013. 30(4): p. 772-780.
 23. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research*, 2014. 42(W1): p. W320-W324.
 24. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution*, 2016. 33(7): p. 1870-1874.
 25. Blom, N., et al., *Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics*, 2004. 4(6): p. 1633-1649.
 26. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics*, 2012. 28(12): p. 1647-1649.
 27. Dilokpimol, A., et al., *Expanding the feruloyl esterase gene family of Aspergillus niger by characterization of a feruloyl esterase, FaeC. New Biotechnology*, 2017. 37: p. 200-209.

28. Knoch, E., et al., *A β -glucuronosyltransferase from *A rabidopsis thaliana* involved in biosynthesis of type II arabinogalactan has a role in cell elongation during seedling growth.* **The Plant Journal**, 2013. 76(6): p. 1016-1029.
29. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* **Nature Methods**, 2012. 9(7): p. 671-675.
30. Britton, H.T.S. and R.A. Robinson, *CXCVIII.—Universal buffer solutions and the dissociation constant of veronal.* **Journal of the Chemical Society (Resumed)**, 1931: p. 1456-1462.
31. Dilokpimol, A., et al., *Recombinant production and characterisation of two related GH5 endo- β -1, 4-mannanases from *Aspergillus nidulans* FGSC A4 showing distinctly different transglycosylation capacity.* **Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics**, 2011. 1814(12): p. 1720-1729.
32. Zeppa, G., L. Conterno, and V. Gerbi, *Determination of organic acids, sugars, diacetyl, and acetoin in cheese by high-performance liquid chromatography.* **Journal of Agricultural and Food Chemistry**, 2001. 49(6): p. 2722-2726.
33. Dilokpimol, A., et al., *Fungal feruloyl esterases: functional validation of genome mining based enzyme discovery including uncharacterized subfamilies.* **New Biotechnology**, 2018. 41: p. 9-14.
34. KOSEKI, T., et al., *An *Aspergillus awamori* acetyltransferase: purification of the enzyme, and cloning and sequencing of the gene.* **Biochemical Journal**, 1997. 326(2): p. 485-490.
35. Komiya, D., et al., *Crystal structure and substrate specificity modification of acetyl xylan esterase from *Aspergillus luchuensis*.* **Applied and Environmental Microbiology**, 2017. 83(20): p. e01251-17.
36. Kroon, P.A., et al., *A modular esterase from *Penicillium funiculosum* which releases ferulic acid from plant cell walls and binds crystalline cellulose contains a carbohydrate binding module.* **European Journal of Biochemistry**, 2000. 267(23): p. 6740-6752.
37. Kühnel, S., et al., *The ferulic acid esterases of *Chrysosporium lucknowense* C1: purification, characterization and their potential application in biorefinery.* **Enzyme and Microbial Technology**, 2012. 50(1): p. 77-85.
38. Topakas, E., et al., *Expression, characterization and structural modelling of a feruloyl esterase from the thermophilic fungus *Myceliophthora thermophila*.* **Applied Microbiology and Biotechnology**, 2012. 94(2): p. 399-411.
39. Altaner, C., et al., *Regioselective deacetylation of cellulose acetates by acetyl xylan esterases of different CE-families.* **Journal of Biotechnology**, 2003. 105(1-2): p. 95-104.
40. Tenkanen, M., et al., *Production, purification and characterization of an esterase liberating phenolic acids from lignocellulose.* **Journal of Biotechnology**, 1991. 18(1-2): p. 69-83.
41. Mansfield, S.D. and J.N. Saddler, *Applications of enzymes to lignocellulose.* 2003: ACS Publications.
42. Puchart, V., et al., *Substrate and positional specificity of feruloyl esterases for monoferuloylated and monoacetylated 4-nitrophenyl glycosides.* **Journal of Biotechnology**, 2007. 127(2): p. 235-243.
43. Koseki, T., et al., *An *Aspergillus oryzae* acetyl xylan esterase: molecular cloning and characteristics of recombinant enzyme expressed in *Pichia pastoris*.* **Journal of Biotechnology**, 2006. 121(3): p. 381-389.
44. Chung, H.-J., et al., *Cloning the gene encoding acetyl xylan esterase from *Aspergillus ficuum* and its expression in *Pichia pastoris*.* **Enzyme and Microbial Technology**, 2002. 31(4): p. 384-391.
45. Pouvreau, L., et al., *Characterization and mode of action of two acetyl xylan esterases from *Chrysosporium lucknowense* C1 active towards acetylated xylans.* **Enzyme and Microbial Technology**, 2011. 49(3): p. 312-320.
46. Huy, N.D., et al., *Cloning and characterization of a novel bifunctional acetyl xylan esterase with carbohydrate binding module from *Phanerochaete chrysosporium*.* **Journal of Bioscience and Bioengineering**, 2013. 115(5): p. 507-513.
47. Crepin, V.F., C.B. Faulds, and I.F. Connerton, *A non-modular type B feruloyl esterase from *Neurospora crassa* exhibits concentration-dependent substrate inhibition.* **Biochemical Journal**, 2003. 370(2): p. 417-427.
48. Topakas, E., C. Vafiadi, and P. Christakopoulos, *Microbial production, characterization and applications of feruloyl esterases.* **Process Biochemistry**, 2007. 42(4): p. 497-509.
49. Downie, R.C., et al., *Assessing European wheat sensitivities to *Parastagonospora nodorum* necrotrophic effectors and fine-mapping the *Snn3-B1* locus conferring sensitivity to the effector *SnTox3*.* **Frontiers in Plant Science**, 2018. 9: p. 881.
50. Couturier, M., et al., **Podospira anserina* hemicellulases potentiate the *Trichoderma reesei* secretome for saccharification of lignocellulosic biomass.* **Applied and Environmental Microbiology**, 2011. 77(1): p.

237-246.

51. Feliu, J., et al., *Insights into exo- and endoglucanase activities of family 6 glycoside hydrolases from *Podospora anserina**. *Applied and Environmental Microbiology*, 2013. 14 (79): p. 4220-4229.
52. Dilokpimol, A., et al., *Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications*. *Biotechnology for Biofuels*, 2016. 9.

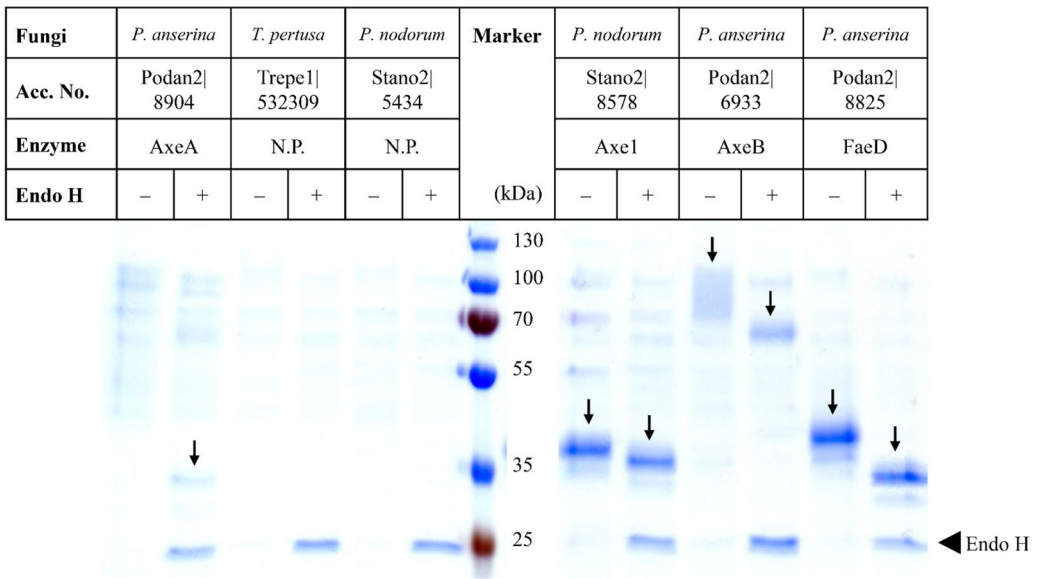
Supplementary material

Suppl. Table 1. Amino acid sequences used for the CE1 phylogeny. The additional sequences are highlighted in green. The outgroup sequences are highlighted in yellow. The selected candidates are highlighted in pink and bold. *Available upon request from author.*

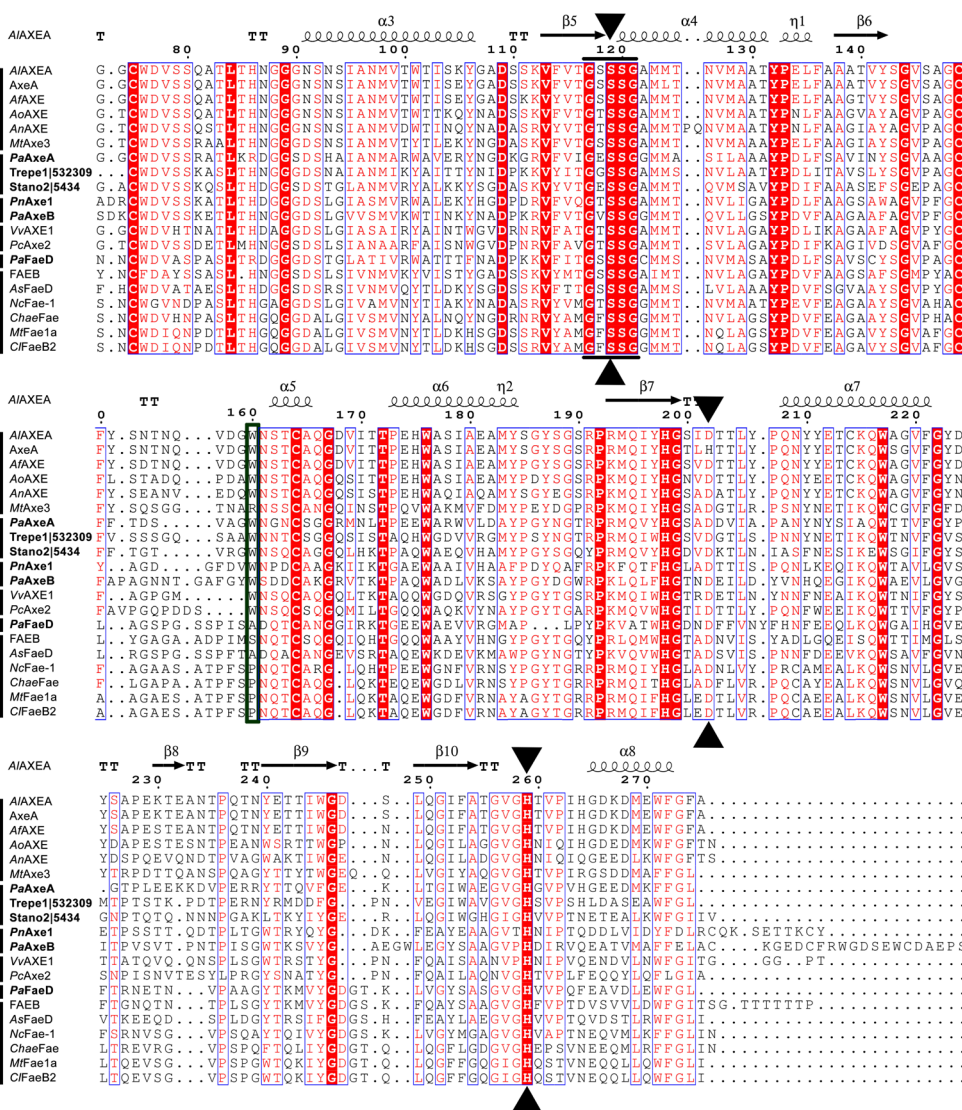
Suppl. Fig. 1 The full phylogenetic tree using total 247 fungal CE1 sequences. The phylogenetic analysis was performed by maximum likelihood (ML) implemented in MEGA7 [24] with 95% partial deletion of gaps and the Poisson correction distance of substitution rates. The main branches/subfamilies were collapsed. Statistical support for phylogenetic grouping was estimated by 500 bootstrap re-samplings, only the bootstrap above 40% were shown on the branches, also of a neighbor-joining (NJ) and minimal evolution (ME) tree using the same dataset (order: ML, NJ, ME). Eight FAEs from subfamily 7 [52] were used as an outgroup. SF, subfamily. The purple highlighted eurotiomycetes, the orange highlighted dothidiomycetes, the red highlighted sordariomycetes, the blue highlighted leotiomycetes, the green highlighted agaricomycotina and the black one is pezizomycetes. *Available upon request from author.*

Suppl. Fig. 2 Recombinant protein production of the selected candidates with (+) and without (-) endoglycosidase H (Endo H) treatment. Black arrows indicate the target bands of candidates and black triangle indicates the endoglycosidase H band. Acc. No., Accession number of the selected candidates. N.P., no production.

Suppl. Fig. 3 Amino acid sequences alignment among the characterized fungal CE1_SF1 and SF2 and the selected candidates. Black triangles indicate (putative) catalytic triad (Ser/Asp/His). The line indicates the G-X-S-X-G motif. Green box indicates Trp160 in *AlAXEA*. Names in bold indicate the selected candidates for this study.



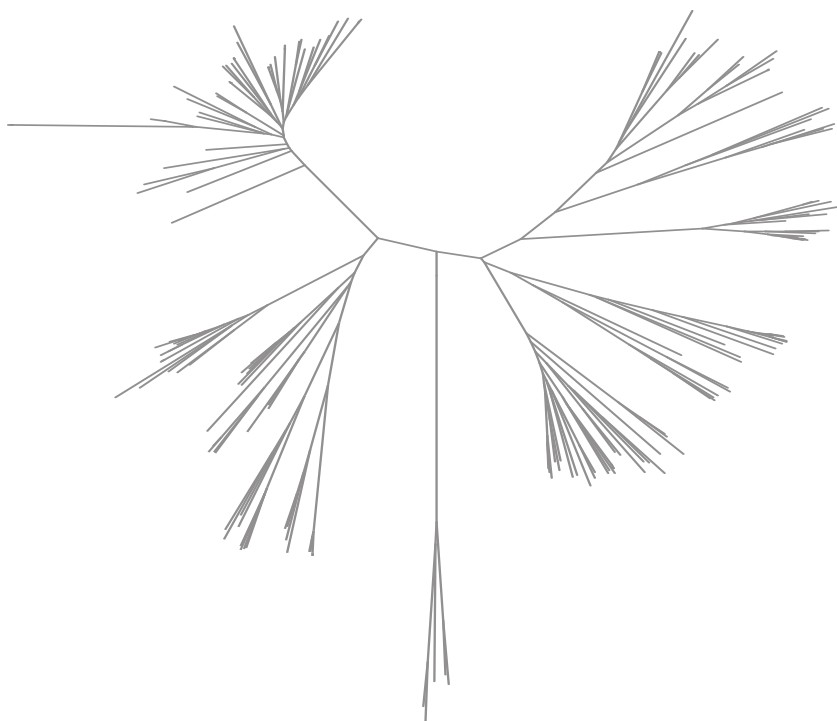
Suppl. Fig. 2 Recombinant protein production of the selected candidates with (+) and without (-) endoglycosidase H (Endo H) treatment. Black arrows indicate the target bands of candidates and black triangle indicates the endoglycosidase H band. Acc. No., Accession number of the selected candidates. N.P., no production.



Suppl. Fig. 3 Amino acid sequences alignment among the characterized fungal CE1_SF1 and SF2 and the selected candidates. Black triangles indicate (putative) catalytic triad (Ser/Asp/His). The line indicates the G-X-S-X-G motif. Green box indicates Trp160 in A1AXEA. Names in bold indicate the selected candidates for this study.

章节三

真菌糖苷酶GH30家族蕴藏多个具有不同多糖特异性的亚家族



Chapter 3

Glycoside Hydrolase family 30 harbors fungal subfamilies with distinct polysaccharide specificities

This chapter was published in *New Biotechnology*

Li, Xinxin, Dimitrios Kouzounis, Mirjam A. Kabel, Ronald P. de Vries, and Adiphol Dilokpimol

Volume 67, 25 March 2022, P32-41

<https://doi.org/10.1016/j.nbt.2021.12.004>

Abstract

Efficient bioconversion of agro-industrial side streams requires a wide range of enzyme activities. Glycoside Hydrolase family 30 (GH30) is a diverse family that contains various catalytic functions and has so far been divided into ten subfamilies (GH30_1-10). In this study, a GH30 phylogenetic tree using over 150 amino acid sequences was constructed. The members of GH30 cluster into four subfamilies and eleven candidates from these subfamilies were selected for biochemical characterization. Novel enzyme activities were identified in GH30. GH30_3 enzymes possess β -(1 \rightarrow 6)-glucanase activity. GH30_5 targets β -(1 \rightarrow 6)-galactan with mainly β -(1 \rightarrow 6)-galactobiohydrolase catalytic behavior. β -(1 \rightarrow 4)-Xylanolytic enzymes belong to GH30_7 targeting β -(1 \rightarrow 4)-xylan with several activities (e.g., xylobiohydrolase, endoxylanase). Additionally, a new fungal subfamily in GH30 was proposed, i.e., GH30_11, which displays β -(1 \rightarrow 6)-galactobiohydrolase. This study confirmed that GH30 fungal subfamilies harbor distinct polysaccharide specificity and have high potential for the production of short (non-digestible) di- and oligosaccharides.

Introduction

Glycoside hydrolases (GHs) are ubiquitously present in nature. They are critical for the efficient degradation of plant biomass in various industrial applications, such as in food and biofuel production, and the pulp and paper industries [1]. GHs are divided into different families based on their amino acid sequence similarities in the Carbohydrate-Active enZYme (CAZy) database [2]. Glycoside Hydrolase family 30 (GH30) is a diverse but understudied family, containing enzymes from bacteria, fungi and animals. Recently, GH30 has gained increased attention because of the discovery of fungal β -(1 \rightarrow 4)-xylobiohydrolases [3, 4]. In addition to this new activity, the GH30 enzyme family also contains β -glucocerebrosidases, β -glucosidases, reducing end β -(1 \rightarrow 4)-exoxyylanases, β -(1 \rightarrow 4)-endoxyylanases, β -(1 \rightarrow 6)-galactanases, β -(1 \rightarrow 6)-glucanases, and β -fucosidases (http://www.cazy.org/GH30_characterized.html) [2, 5].

GH30 was previously divided into ten subfamilies (GH30_1-10) based on phylogenetic analysis [5-7] with fungal candidates in only GH30_3, 5 and 7 (**Figure 1**, **Table 1**). In GH30_3, four fungal enzymes have been previously characterized, all showing β -(1 \rightarrow 6)-glucanase activity [8-12]. In GH30_5, only two fungal enzymes have been characterized. These β -(1 \rightarrow 6)-galactanases hydrolyze between D-galactosyl moieties in type II arabinogalactan [13, 14]. In contrast to GH30_3 and 5, which showed similar activities within the subfamily, the nine characterized enzymes from GH30_7 hydrolyze β -(1 \rightarrow 4)-xylan with diverse activities (**Table 1**). These activities include β -(1 \rightarrow 4)-endoxyylanases, β -(1 \rightarrow 4)-xylobiohydrolases and reducing end β -(1 \rightarrow 4)-exoxyylanases.

Based on current knowledge, fungal GH30 enzymes could be used to depolymerize different polysaccharides and generate 'short' non-digestible di- and oligosaccharides, e.g., β -(1 \rightarrow 6)-glucobiose/glucooligosaccharides, β -(1 \rightarrow 6)-galactobiose/galactooligosaccharides, and β -(1 \rightarrow 4)-xylobiose/xylooligosaccharides, which are of particularly interest in food and feed industries [15, 16]. Non-digestible oligosaccharides (NDOs) have been reported to exhibit prebiotic potential by selectively stimulating the growth and activity of beneficial bacteria in the colon and promote the health of the host [17]. However, these enzymes can also be used in other biotechnology applications. For example, β -(1 \rightarrow 6)-glucanases from GH30_3 could be included in antifungal products due to their fungal cell wall degrading capabilities [18, 19], while β -(1 \rightarrow 4)-xylanolytic enzymes from GH30_7 enzymes can be applied for biomass saccharification for the production of biofuels and fine chemicals [20].

In this study, to verify the functional specificity of these identified fungal GH30 subfamilies and deepen understanding of the fungal members in GH30, a GH30 phylogenetic tree was constructed using 161 amino acid sequences from bacterial, fungal and animal species, from which eleven candidates covering different fungal subfamilies were selected for biochemical characterization using polysaccharides and crude plant biomass. This revealed a new fungal subfamily in GH30. After discussion with curators from the CAZy Database, this is now established as GH30_11, which displays β -(1 \rightarrow 6)-galactobiohydrolase activity. In addition, biochemical characterization revealed that the enzymes from different GH30 subfamilies exhibited distinct substrate specificities. Furthermore, to investigate the catalytic mechanism of GH30 fungal enzymes, the potential residues involved in substrate-protein interactions were analyzed by homology modelling, from which some possible amino acid residues affecting substrate specificity were highlighted. In conclusion, this study facilitates the industrial use of fungal GH30 enzymes, as the knowledge of different fungal subfamilies will help to select optimal candidates for the desired application, be that for the production of specific short NDOs or for other biorefinery process.

Materials and methods

Materials

β -(1 \rightarrow 6)-Glucan (Pustulan) was from Carbosynth Ltd. (Berkshire, UK). Beech wood xylan (BeWX) was from Carl Roth GmbH + Co. KG (Karlsruhe, Germany). Wheat bran (WB) was from Wageningen Mill (Wageningen, the Netherlands). β -(1 \rightarrow 3),(1 \rightarrow 4)-Glucan (from Barley), insoluble wheat arabinoxylan (WAX, P-WAXYI, from wheat flour), α -L-arabinofuranosidase (ABF, E-AFASE, GH51 from *Aspergillus niger*), α -glucuronosidase (AGU, E-AGUBS, GH67 from *Geobacillus stearothermophilus*) and oligosaccharides were from Megazyme (Bray, Ireland). β -(1 \rightarrow 4)-Glucan (Avicel[®]PH-101), β -(1 \rightarrow 3)-glucan (from *Euglena gracilis*), β -(1 \rightarrow 3),(1 \rightarrow 6)-glucan (Laminarin), Larch wood arabinogalactan II (LWAG-II), Gum Arabic arabinogalactan II (GAAG-II), pectin (from potato, apple and esterified citrus), 4-*O*-methyl-D-glucurono-D-xylan (MGX) and other chemicals were from Sigma-Aldrich (Merck KGaA, Darmstadt, Germany).

Bioinformatics and homology modelling

The fungal sequences from JGI Mycocosm (<https://mycocosm.jgi.doe.gov/mycocosm/home>), the sequences from characterized GH30 enzymes in CAZy database (<http://www.cazy.org/>), the additional sequences of bacteria and animals obtained by BLASTP from the non-redundant sequence database (<https://www.ncbi.nlm.nih.gov/>) were used for phylogenetic analysis (a list of fungal species is shown in **Suppl. Table 1**). The secretory signal peptides were detected using SignalP 4.1 (<http://www.cbs.dtu.dk/services/SignalP/>). The amino acid sequences without predicted signal peptides were aligned using MAFFT (<https://www.ebi.ac.uk/Tools/msa/mafft/>) and visualized using Easy Sequencing in Postscript 3.0 (<http://escript.ibcp.fr/EScript/EScript/index.php>) [21, 22]. The phylogenetic analysis was performed using maximum likelihood (ML) implemented in the Molecular Evolutionary Genetic Analysis software version 7 (MEGA7) as described previously [23] with 95% partial deletion and the Poisson correction distance of substitution rates [24]. Statistical support for phylogenetic grouping was estimated by 500 bootstrap re-samplings. Only bootstrap values above 50% are shown in the tree. Homology models were generated using the Homology detection and structure prediction by HMM-HMM comparison (HHpred, <https://toolkit.tuebingen.mpg.de/tools/hhpred>) [25]. The structure of *TeXyn30B* from *Talaromyces cellulolyticus* (PDB ID: 6KRN) [26] was used as a template. The quality of the models were validated as described previously [27]. To compare the putative catalytic residues, all models were superimposed on *TeXyn30B* complexed with glucuronyl-xylobiose (GlcA-X2) ligand (PDB ID: 6KRN) [26] and visualized by PyMOL 2.3.5 (Schrödinger, Inc., New York, NY). In addition, the structure of *Erwinia chrysanthemi EcXyn30A* complexed with glucuronyl-xylotriose (GlcA-X3) ligand (PDB ID: 2Y24) [28] was also used to define the putative subsites (-1 to -3) in the catalytic domain.

Cloning, protein production and purification of the selected candidates

The selected genes (**Figure 1, Table 1**) without predicted signal peptides and introns were codon optimized and synthesized into pPicZaA plasmid (Genscript Biotech, Leiden, the Netherlands), which was then transformed into *Escherichia coli* DH5 α (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA) and subsequently transformed into *Pichia pastoris* strain X-33 (Invitrogen) according to the manufacturer's recommendation. The positive colonies were selected as described earlier [23].

The selected *P. pastoris* transformants were grown and induced for the production of enzyme as described previously [29]. Culture supernatants were harvested (8000 \times g, 4°C, 20 min). One portion was filtered (0.22 μ m; Merck), aliquoted and stored at -20°C (as crude enzyme). The other part was purified by ÄKTA FPLC System (GE Life Sciences, Uppsala, Sweden). Crude enzymes were loaded onto a HisTrap FF 1 mL column (Cytiva, Marlborough, USA) equilibrated with 20 mM HEPES, 0.4 M NaCl, 20 mM imidazole, pH 7.5, and eluted using a linear gradient of 22-400 mM imidazole in buffer mentioned above at a flow rate of 1.0 mL/min. Fractions containing enzyme were collected, concentrated and buffer-exchanged to 20 mM HEPES (pH 7.0) using 10 kDa cut-off ultrafiltration units Amicon (Millipore, Bedford, MA). All purification steps were performed at 4 °C

Table 1 Substrate specificity of characterized enzymes and selected candidates from fungal subfamilies of GH30.

Fungal species	Gene name	Enzyme name	Accession no.	Known activity	Linkage	Mainly active polysaccharides	SF	References
<i>Aspergillus fumigatus</i>	neg1	Neg1	EAL85472.1	Glucanase	β -(1→6)	β -(1→6)-glucan	3	[8]
<i>Trichoderma harzianum</i>	N.A.	BGN16.3	CAC80492.1	Glucanase	β -(1→6)	β -(1→6)-glucan	3	[9]
<i>Neurospora crassa</i>	neg1	Neg1	BAB91213.1	Glucanase	β -(1→6)	β -(1→6)-glucan	3	[10]
<i>Penicillium subrubescens</i>	engA	P3EngA	jgi Pensub1 3158	Glucanase	β -(1→6)	β -(1→6)-glucan	3	This study
<i>Aspergillus niger</i>	engA	AnEngA	jgi Aspni 8738	Glucanase	β -(1→6)	β -(1→6)-glucan	3	This study
<i>Talaromyces atrovirens</i>	engA	TzEngA	XP020122703.1	Glucanase	β -(1→6)	β -(1→6)-glucan	3	This study
<i>Talaromyces emersonii</i> ^a	engA	TzEngA	XP013323429.1	Glucanase	β -(1→6)	β -(1→6)-glucan	3	This study
<i>Lentinula edodes</i>	lep30a	LePus30A	BAK52530.1	Glucanase	β -(1→6)	β -(1→6)-glucan	3	[11]
<i>Stereum hirsutum</i>	gbhA	StGbhA	jgi Stehi 99993	Galactobioreductase	β -(1→6)	Type II arabinogalactan	11	This study
<i>Trametes versicolor</i>	gbh1	TrGbh1	jgi Trave 66957	Galactobioreductase	β -(1→6)	Type II arabinogalactan	11	This study
<i>Neurospora crassa</i>	Nc6GAL	Nc6GAL	XP330352.1	Galactamase	β -(1→6)	Type II arabinogalactan	5	[14]
<i>Penicillium subrubescens</i>	gbhA	P3GbhA	jgi Pensub1 11412	Galactobioreductase	β -(1→6)	Type II arabinogalactan	5	This study
<i>Trichoderma viride</i>	Tv6GAL	Tv6GAL	BAC84995.1	Galactamase	β -(1→6)	Type II arabinogalactan	5	[35]
<i>Talaromyces cellulolyticus</i>	xyn30C	TzXyn30C	GAM40414.1	Endoxylanase	β -(1→4)	Glucuro- / arabinoxylan	7	[54]
<i>Bispora sp. MEY-1</i>	xy1D	XY1D	ADG62369.1	Endoxylanase	β -(1→4)	Glucuro- / arabinoxylan	7	[36]
<i>Penicillium purpogenum</i> ^b	xynC	XynC	AKH40280.1	Endoxylanase	β -(1→4)	Glucuro- / arabinoxylan	7	[37]
<i>Trichoderma reesei</i>	XYN VI	XYN VI	EGR45006.1	Endoxylanase	β -(1→4)	Glucuronoxylan	7	[38]
<i>Penicillium subrubescens</i>	ex1A	P3Ex1A	jgi Pensub1 7918	Endoxylanase	β -(1→4)	Glucuronoxylan	7	This study
<i>Trichoderma reesei</i>	xyn4	XYN IV	AAP64786.1	Endoxylanase / reducing end exoxylanases	β -(1→4)	Rhodymenan / glucuronoxylan	7	[39]
<i>Talaromyces cellulolyticus</i>	xyn30A	TzXyn30A	GAM43270.1	reducing end exoxylanases	β -(1→4)	Glucuro- / arabinoxylan	7	[40]
<i>Acremonium alticola</i>	AaXyn30A	AaXyn30A	N.A.	Xylobiodydrolase	β -(1→4)	Rhodymenan / glucuronoxylan	7	[3]
<i>Thermothelomyces thermophilus</i> ^c	xbhA	TrXbhA	jgi Spoh2 101800	Xylobiodydrolase	β -(1→4)	Glucuronoxylan	7	This study
<i>Talaromyces emersonii</i>	xbhA	TzXbhA	KKAL17994.1	Xylobiodydrolase	β -(1→4)	Glucuronoxylan	7	This study
<i>Talaromyces cellulolyticus</i>	xyn30B	TzXyn30B	GAM36763	Endoxylanase / xylobiodydrolase	β -(1→4)	Glucuronoxylan	7	[41]
<i>Thermothelomyces thermophilus</i>	ex1A	TrEx1A ^d	XP003660270.1	Endoxylanase / xylobiodydrolase	β -(1→4)	Glucuronoxylan	7	This study

N.A., not available; SF, subfamily.

a, The homotypic synonym: *Rasamsonia emersonii*.

b, The homotypic synonym: *Talaromyces purpogenum*.

c, The homotypic synonym: *Myceliophthora thermophila*.

d, Formerly named *TrXyn30A* [4].



[27]. The concentration of purified protein was evaluated as described earlier [23].

Enzyme activity assays

Activity screening using crude enzymes

The activity assays of crude enzymes were performed in 600 μ L reaction mixtures containing 500 μ L 1% (w/v) of the substrates (**Figure 2, Suppl. Table 2**) in 50 mM NaOAc buffer (pH 5.5) and 100 μ L crude enzymes at 30°C, 110 rpm for 16 h. The culture supernatant from *P. pastoris* harboring pPicZaA plasmid without insertion was used as a negative control for crude enzymes.

Activity assays using purified enzymes

The activity of the purified enzymes was assayed using 3 mg/mL of the selected polysaccharides in 50 mM NaOAc (pH 5.5) incubated with 6 μ g/mL purified enzyme. HEPES buffer was used as a negative control for purified enzymes. The reaction was performed at 30°C, 110 rpm, for up to 16 h, and aliquots of 500 μ L were removed from the mixtures at different time points (5 min, 15 min, 30 min, 45min, 1 h, 2 h, 6 h and 16 h) for time course analysis.

To validate the activity of GH30_3 enzymes towards β -glucans with different linkages, 3 mg/mL β -glucans in 50 mM NaOAc (pH 5.5) were incubated with 6 or 60 μ g/mL purified enzyme, respectively, at 30°C for 16 h.

To investigate the effect of xylan substitutions on GH30_7 enzyme activity, MGX and WAX were used as representative substrates for 4-*O*-methylglucuronic acid (MeGlcA) substituted and arabinofuranose substituted xylans, respectively [30]. 3 mg/mL MGX and WAX were incubated with 0.17 mg/mL AGU and ABF, respectively, in 50 mM NaOAc (pH 5.5) and 0.02% NaN₃ at 30°C, 110 rpm, for 72 h. The reaction was stopped by heating at 95°C for 10 min. These are referred to as pre-treated substrates. For the activity assay, the reaction mixtures containing 550 μ L 3 mg/mL MGX, WAX or pre-treated substrates and 50 μ L purified GH30_7 enzymes (equal to approximately 3 μ g protein) were incubated under the above-mentioned conditions.

Mono- and oligosaccharide analysis

All mentioned reactions were stopped by heat inactivation at 95°C for 10 min and used for HPAEC-PAD analysis. The reaction mixtures were centrifuged (14,000 \times g, 4°C, 15 min) and the supernatant was diluted 10-fold in Ultrapure water (18.2 M Ω cm; Elga PURELAB flex, Marlow, UK) prior to the analysis. The mono- and oligosaccharide release was quantified using HPAEC-PAD (Dionex ICS-5000+ system) (Thermo Fisher Scientific, Sunnyvale, CA) equipped with a CarboPac PA1 (250 mm \times 4 mm i.d.) column (Thermo Fisher Scientific) with different profiles according to [27] and [31], respectively. 5-250 μ M different monosaccharides (D-glucose, D-galactose, L-arabinose, D-xylose) and disaccharides (cellobiose, β -(1 \rightarrow 6)-glucobiose, β -(1 \rightarrow 6)-galactobiose, β -(1 \rightarrow 4)-xylobiose) were used as standards for identification and quantification. Total sugar composition of the used substrates was determined according to [32].

The released glucuronic acid was quantified using HPAEC-PAD in the same way as for monitoring monosaccharide release. The chromatographic separation was performed with a multi-step gradient according to [27] using the following procedure: 0–20 min: 18 mM NaOH, 20–30 min: 0–40 mM NaOH and 0–400 mM sodium acetate, 30–35 min: 40–100 mM NaOH and 400 mM to 1 M sodium acetate, 35–40 min: 100 mM NaOH and 1 M to 0 M sodium acetate, 10 min: 18 mM NaOH (20 °C; flow rate: 0.30 mL/min).

Results and Discussion

Fungal GH30 enzymes belong to four subfamilies

Phylogenetic analysis of GH30 updated the fungal subfamilies included in the CAZy database (**Figure 1A, Suppl. Fig. 1**) (http://www.cazy.org/GH30_characterized.html). Fungal members clustered into

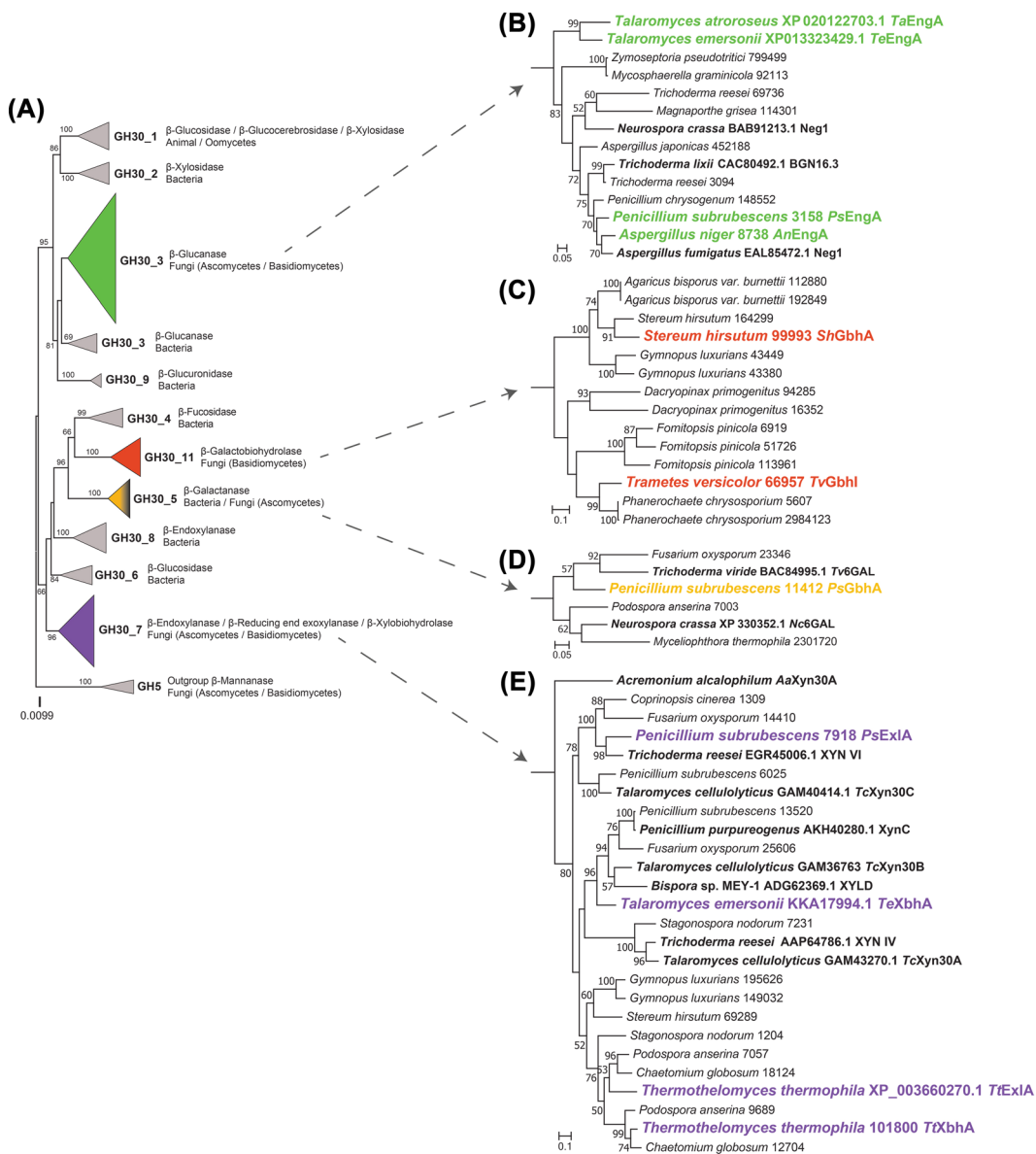


Figure 1 Phylogenetic relationship among GH30 members from fungi, bacteria, and animals based on their amino acid sequences. The phylogenetic analysis was performed as previously reported [23]. The main subfamilies were collapsed. Statistical support for phylogenetic grouping was estimated by 500 bootstrap re-samplings, only the bootstrap above 50% were shown on the clades. Five mannanases from GH5 were used as an outgroup. GH30_10 containing bacterial xylobiohydrolases was identified by Crooks, Bechle [6] which is not shown in the current GH30 tree. (A) Whole tree; (B) GH30_3; (C) GH30_11; (D) GH30_5; (E) GH30_7. Selected candidates are highlighted and the characterized enzymes are in boldface. The full phylogenetic tree can be found in **Suppl. Fig. 1**.

the previously defined GH30_3, 5 and 7 and also into the newly classified GH30_11 subfamily (this work). In the GH30 phylogenetic tree (*the order of description is based on the location of subfamilies in the tree*), GH30_3 members were further divided into two main clades based on different kingdoms, i.e., fungi and bacteria. The fungal clade was predicted to possess β -(1 \rightarrow 6)-glucanase (ENG) activity based on earlier characterization [8-11] (**Figure 1A, B**). The fungal GH30_11 shares the same node with the bacterial GH30_4, which suggests that they were more related to each other than to the other subfamilies. However, the sequence alignment showed low amino acid sequence identity (<28%) between members from GH30_11 and GH30_4, indicating that they might differ in catalytic function. GH30_11 contained only sequences from basidiomycetes (**Figure 1C**), but no characterized enzyme from this subfamily has been reported until now. A previous study found two β -D-fucosidases in GH30_4 [33]. Similar to GH30_3, members of GH30_5 were classified into two clades (fungi and bacteria). However, the fungal clade of GH30_5 only contained sequences from ascomycetes (**Figure 1D**). The fungal and bacterial GH30_5 showing similar activity (β -(1 \rightarrow 6)-galactanase) shared more than 50% amino acid sequence identity [13, 14, 34]. GH30_7 contained sequences from both ascomycetes and basidiomycetes (**Figure 1E**) and the characterized enzymes showed several xylanolytic activities (**Table 1**).

To evaluate different fungal subfamilies systematically and compare their activities and substrate specificity, eleven candidates from these four subfamilies were selected for biochemical characterization: four from GH30_3, two from the newly established GH30_11, one from GH30_5, and four from GH30_7.

Different fungal subfamilies target different substrates

The activity of culture supernatants was first screened towards several substrates (**Figure 2, Suppl. Table 2**): β -glucans with different linkages, arabinogalactan II (AG-II), and xylan and pectin from different sources. The screening results showed that selected candidates exhibited different linkage specificities, with GH30_3 candidates acting on D-glucosyl linkages, GH30_11 and GH30_5 candidates hydrolyzing D-galactosyl linkages, and GH30_7 candidates breaking D-xylosyl linkages. The corresponding mono- and di-saccharides released from hydrolyzing the tested substrates by different candidates are shown in **Figure 2**.

GH30_3 candidates were active towards various substrates, but showed a preference for β -(1 \rightarrow 6)-glucan (**Figure 2A**). Their major disaccharide product was β -(1 \rightarrow 6)-glucobiose, whereas minor products, e.g., β -(1 \rightarrow 4)-glucobiose, were released from other substrates. The release of these saccharides from xylans and pectins were presumably due to the presence of small amounts of glucan or cellulose in those samples.

GH30_11 and GH30_5 candidates were mainly active against AG-II, especially Larch wood arabinogalactan II (LWAG-II), and released more β -(1 \rightarrow 6)-galactobiose than monomeric D-galactose (**Figure 2B, 2C**). AG-II consists of a β -(1 \rightarrow 3)-galactan backbone substituted with β -(1 \rightarrow 6)-galactooligosaccharides, whereas the galactans in pectin consist mainly of β -(1 \rightarrow 4)-linkages [42-44]. The results suggested that GH30_11 and GH30_5 enzymes target β -(1 \rightarrow 6)-galactan, which agreed with the previous reports of GH30_5 [13, 14]. In addition, *PsGbhA* also released a small amount of D-galactose from wheat bran (WB, **Suppl. Table 2**), which might be due to the presence of AG-II in WB [45].

GH30_7 were only active on xylan substrates, i.e., Beech wood xylan (BeWX) and WB, and showed a high preference for BeWX (**Figure 2D**). The candidates from GH30_7 all released xylobiose as the major product from BeWX, while the production of other xylooligosaccharides (XOS) was enzyme dependent. This is discussed in more detail below.

Based on the initial screening results of the crude enzymes, β -(1 \rightarrow 6)-glucan, LWAG-II and BeWX were selected as the main polysaccharides for more detailed characterization of candidate enzymes from GH30_3, GH30_11 & GH30_5, and GH30_7, respectively.

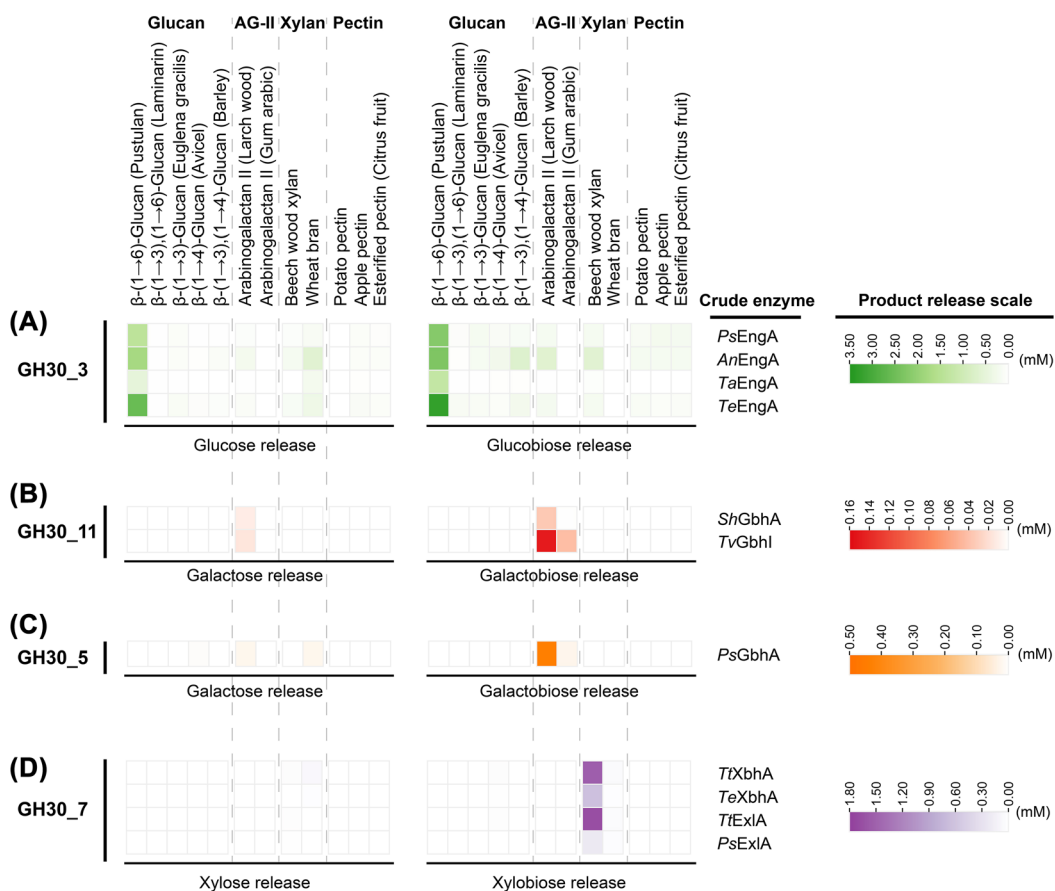


Figure 2 Mono- and di-saccharides released from different substrates by crude GH30 enzymes. Hydrolysis was performed at 30°C for 16 h. Saccharide release was detected by HPAEC-PAD. AG-II, arabinogalactan type II. (A)-(D) the type and amount of specific products released from different substrates by crude enzymes of each subfamily. The amount of glucobiose, galactobiose and xylobiose released was calculated based on cellobiose, β -(1 \rightarrow 6)-galactobiose, and β -(1 \rightarrow 4)-xylobiose standard curves, respectively. All assays were performed in duplicate.

Purified GH30_3 enzymes possess β -(1 \rightarrow 6)-glucanase activity

The purified GH30_3 enzymes were not only tested on β -(1 \rightarrow 6)-glucan as a substrate, but also on four other β -glucans with different linkages to further explore their linkage specificity. Unlike the crude GH30_3 candidates, which could hydrolyze β -glucans with different linkages, the purified GH30_3 candidates were solely active towards β -(1 \rightarrow 6)-glucan. The main product was β -(1 \rightarrow 6)-glucobiose followed by D-glucose and β -(1 \rightarrow 6)-glucooligosaccharides (**Figure 3A**). The release of β -(1 \rightarrow 6)-glucobiose and D-glucose was also detected from β -(1 \rightarrow 3),(1 \rightarrow 6)-glucan after increasing the enzyme concentration 10-fold (data not shown). This showed that GH30_3 enzymes only have trace or unspecific activity towards other linkages. These results agreed with previous studies showing that *T. harzianum* BGN16.3 from ascomycete clade [9] and *L. edodes* LePus30A from basidiomycete clade [11] were mainly active towards β -(1 \rightarrow 6)-glucan and had much lower activity towards β -(1 \rightarrow 3),(1 \rightarrow 6)-glucan. *TeEngA* was the most active GH30_3 enzyme tested here, and considerable release of β -(1 \rightarrow 6)-glucobiose was observed within 1 h, which was then further hydrolyzed to D-glucose (**Figure 3B, 3C**).

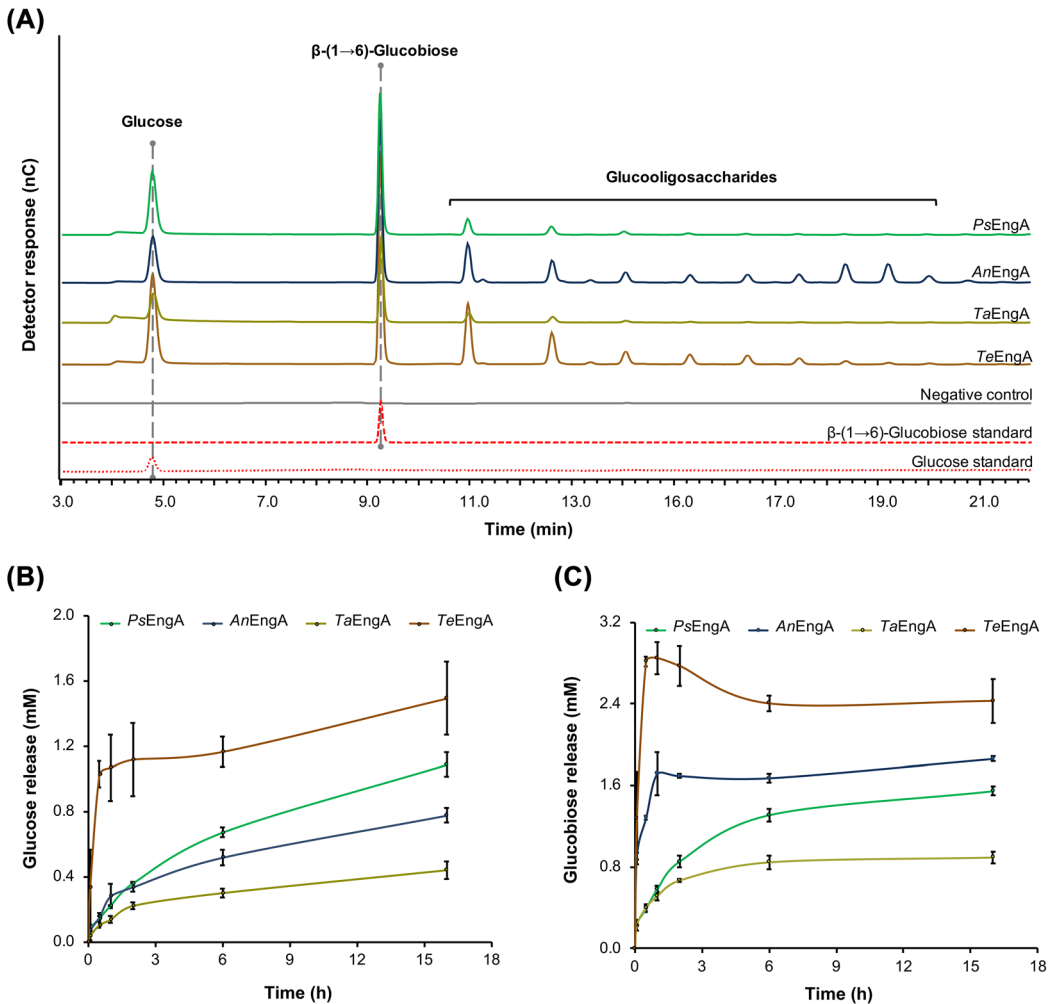


Figure 3 Mono- and oligo-saccharides released from β -(1 \rightarrow 6)-glucan by purified GH30_3 enzymes. (A) HPAEC-PAD profile of saccharide released in 16 h; (B) the amount of glucose released from β -(1 \rightarrow 6)-glucan; (C) the amount of β -(1 \rightarrow 6)-glucobiose released from β -(1 \rightarrow 6)-glucan. Hydrolysis was performed in NaOAc (pH 5.5) at 30°C for up to 16 h. All assays were performed in duplicate.

Purified GH30_11 and GH30_5 enzymes possess β -(1 \rightarrow 6)-galactobiohydrolase activity

LWAG-II was selected as the main substrate for the purified GH30_11 and GH30_5 enzymes, which released only β -(1 \rightarrow 6)-galactobiose from LWAG-II (**Figure 4A**). Note that low amounts of D-galactose and L-arabinose were already present in LWAG-II (**Figure 4B**), and no increased levels of these monosaccharides were detected after 16 h incubation. Hence, a specific β -(1 \rightarrow 6)-galactobiohydrolase activity was concluded for these enzymes. For GH30_5, two fungal enzymes, i.e., *T. viride* Tv6GAL [35] and *N. crassa* Nc6GAL [14] have been previously characterized. Both of these enzymes showed endo β -(1 \rightarrow 6)-galactanase activity, of which Tv6Gal released β -(1 \rightarrow 6)-galactobiose as the major product from pre-treated arabinogalactan-protein from radish, while Nc6GAL specifically targeted β -(1 \rightarrow 6)-galactooligosaccharides and released mainly β -(1 \rightarrow 6)-galactobiose. No fungal GH30_11 enzyme has been previously characterized. Based on these results, β -(1 \rightarrow 6)-galactobiohydrolase activity should be considered in the evaluation of fungal GH30_11 and GH30_5.

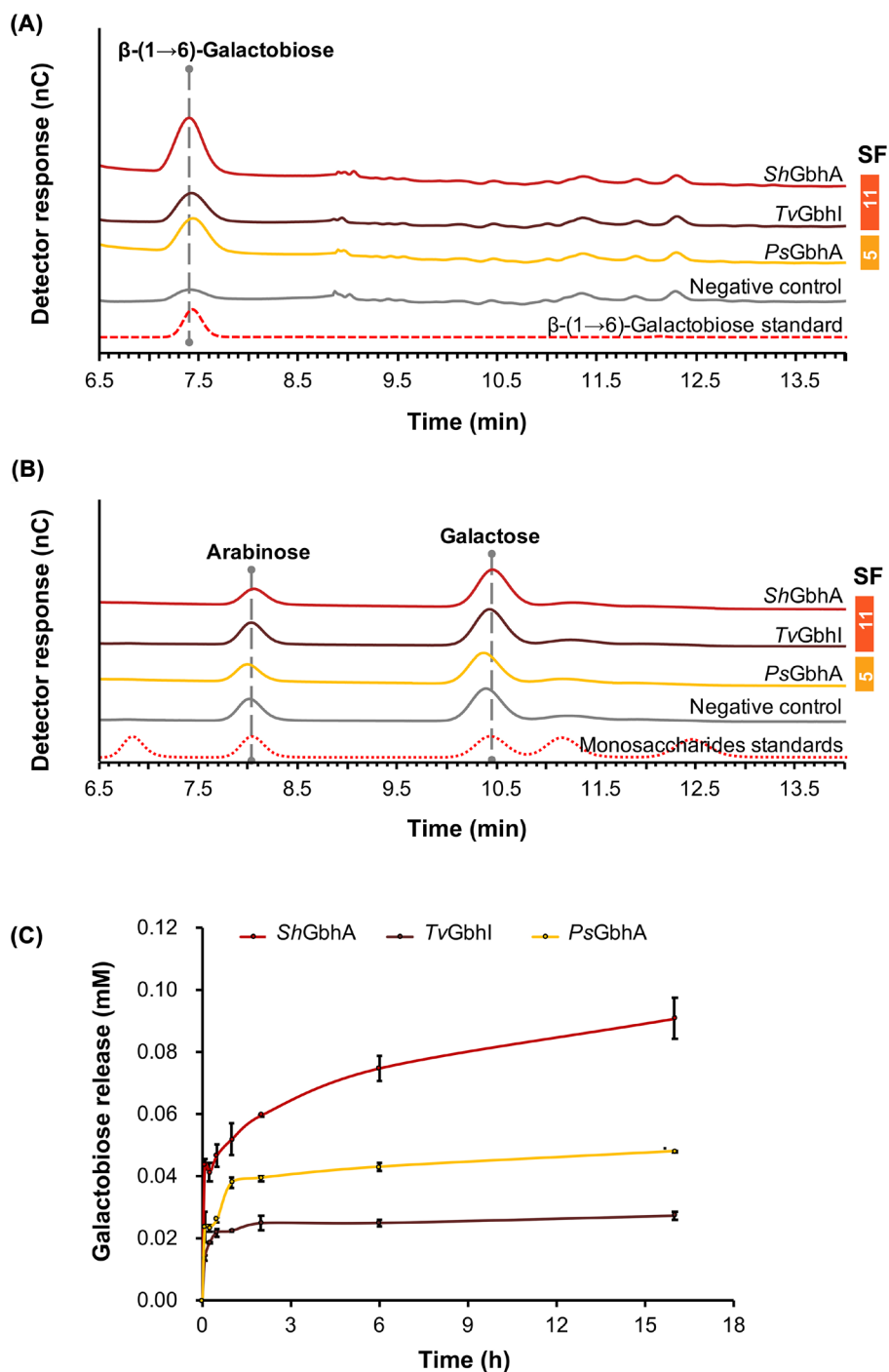


Figure 4 Mono- and oligo-saccharides released from Larch wood arabinogalactan II (LWAG-II) by purified GH30_11 and GH30_5 enzymes. (A) HPAEC-PAD profile of galactooligosaccharides released from the hydrolysis of LWAG-II. Hydrolysis was performed at 30°C for 16 h. (B) Monosaccharides released from the hydrolysis of LWAG-II; (C) the amount of β -(1 \rightarrow 6)-galactobiose released. Hydrolysis was performed in NaOAc (pH 5.5) at 30°C for up to 16 h. All assays were performed in duplicate. SF, subfamily.

Of the GH30_11 and GH30_5 enzymes tested, *ShGbhA* was the most active and released 2- and 4-fold higher amounts of β -(1 \rightarrow 6)-galactobiose than *TvGbhI* and *PsGbhA*, respectively (**Figure 4C**). In addition, *ShGbhA* continuously released β -(1 \rightarrow 6)-galactobiose over a period of 16 h, whereas *PsGbhA* and *TvGbhI* reached a saturation level after 1 h. This suggests different substrate specificity or stability among these enzymes. GH30_11 and GH30_5 enzymes (**Figure 4C**) showed a limited ability to hydrolyze LWAG-II. This might be related to the structure of LWAG-II, which consists of a β -(1 \rightarrow 3)-galactan backbone with β -(1 \rightarrow 6)-galactooligosaccharide side chains. The side chains are highly variable in length and some are further substituted by arabinofuranose, which could affect the enzyme activity [44].

Purified GH30_7 enzymes possess xyloanalytic activity with different product profiles

BeWX degradation revealed various activities of GH30_7 enzymes

BeWX was selected as the substrate for purified GH30_7 enzymes. *TtXbhA* and *TeXbhA* released almost exclusively xylobiose from BeWX, while *PsExlA* released different MeGlcA substituted XOS (**Figure 5A**), indicating that they possess xylobiohydrolase and endoxyylanase activities, respectively. The product pattern of *TtXbhA* and *TeXbhA* in this study differed slightly from the previously reported xylobiohydrolases (e.g., *TtXyn30A*, *TcXyn30B*, *AaXyn30A*), although they all released xylobiose as the major product from BeWX after a long time incubation. *TtXbhA* and *TeXbhA* could release high-purity xylobiose from BeWX, whereas other oligosaccharides (e.g., X4, X5) and a set of MeGlcA substituted XOS were also detected from the BeWX hydrolysis with *TtXyn30A*, *TcXyn30B*, and *AaXyn30A* [3, 4, 41]. *TtXyn30A* and *TcXyn30B* were reported to be bifunctional MeGlcA appendage-dependent xylobiohydrolases / endoxyylanases [3, 41], while *AaXyn30A* is a more strict fungal xylobiohydrolase [4]. *AaXyn30A* also shows slight endoxyylanase activity, which was likely to be the result of excessive enzyme loading in the experimental setup [4, 6].

PsExlA showed a similar hydrolytic product pattern to *T. reesei* XYN VI [38], both of which resembled the bacterial endoxyylanases from GH30_8 [28, 46-48]. *TtExlA* predominantly released xylobiose, but also different (MeGlcA substituted) XOS (**Figure 5A**), similar to the earlier report [4], which shows that it is an endoxyylanase with xylobiohydrolase activity. Among the four tested enzymes, *TtXbhA* was the most suitable for xylobiose production as it released the highest amount of xylobiose (**Figure 5B**).

Considering the different product profiles from BeWX (glucuronoxylan) hydrolysis, the effect of MeGlcA substitution on GH30_7 activity was investigated. Given the low level of MeGlcA substitutions in BeWX (mole ratio MeGlcA:Xyl = 1:15, [49]), 4-*O*-methyl-D-glucurono-D-xylan (MGX; mole ratio MeGlcA:Xyl = 1:5 [50]) was used instead of BeWX in this experiment. GH67 α -glucuronosidase (AGU) from *Geobacillus stearothermophilus* could partially hydrolyze the α -1,2-glycosidic bond between MeGlcA and terminal non-reducing D-xylosyl residues of xylooligosaccharides and xylan [51]. MGX contains approximately 0.1 mg total uronic acid per mg polysaccharide. Only around 8% MeGlcA (of the total uronic acid) was released from MGX by GH67 AGU. The comparison of the amount of xylobiose released from AGU-treated and untreated MGX revealed that the activity of *TeXbhA* was most affected. For this GH30_7 enzyme, a decrease in the yield of xylobiose was observed with the AGU-treated MGX, while the activity of the other enzymes was only slightly or not affected (**Figure 5D**). It is not understood how such a minor decrease in the MeGlcA substitution could decrease the yield of xylobiose for *TeXbhA*. Further studies are required to address the fungal GH30_7 functional requirements for the MeGlcA and the yield of xylobiose.

WAX degradation confirmed the BeWX preference of GH30_7 enzymes

To validate the activity of GH30_7 enzymes towards different types of xylan, they were also analyzed using wheat arabinoxylan (WAX) as a substrate. *TtXbhA* released around 10-fold less

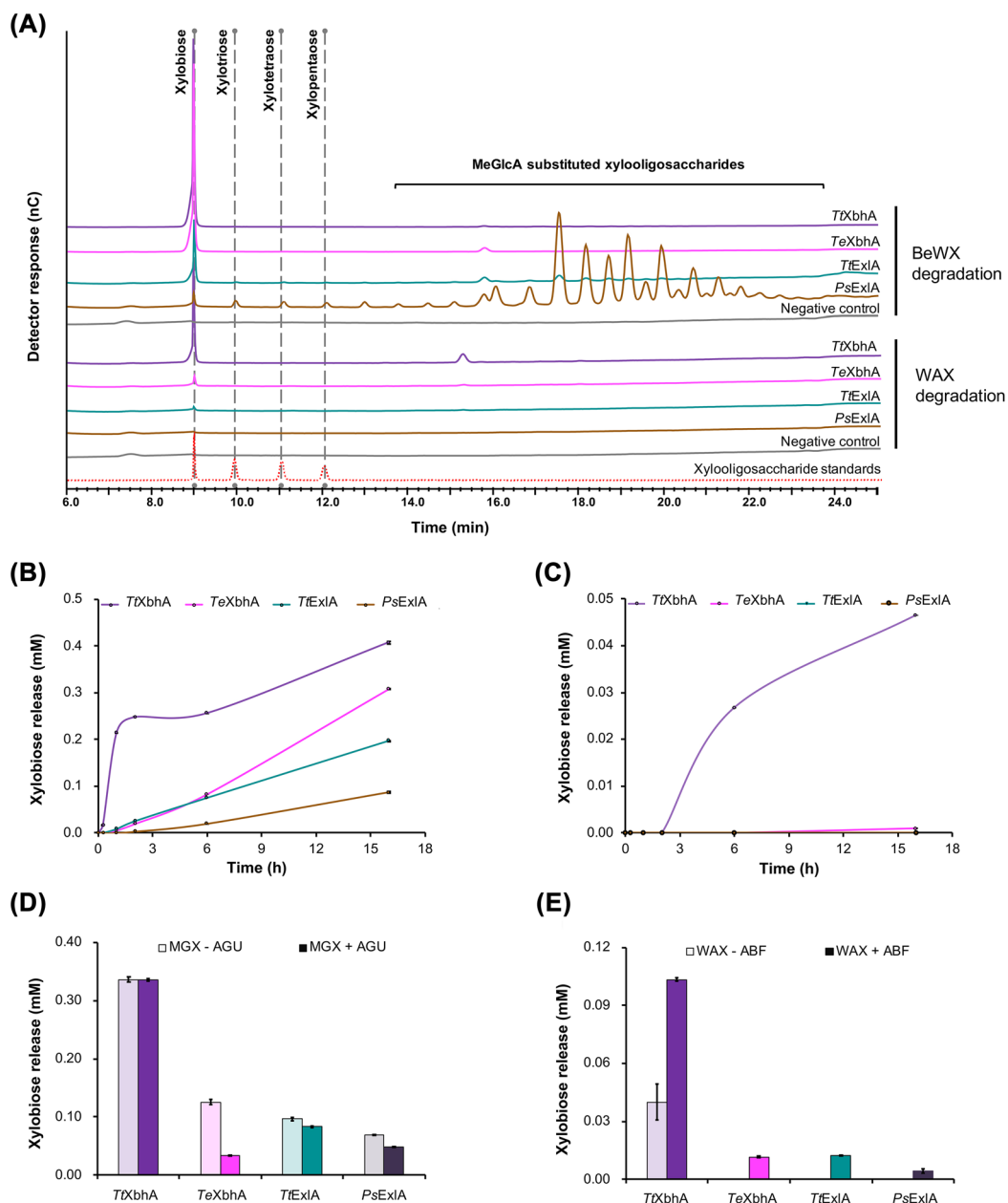


Figure 5 Mono- and oligo-saccharides released from Beech wood xylan (BeWX) and wheat arabinoxylan (WAX) by purified GH30_7 enzymes. (A) HPAEC-PAD profile from the hydrolysis of BeWX (top) and WAX (bottom). Hydrolysis was performed at 30°C for 16 h. (B) The amount of β -1 \rightarrow 4-xylobiose released from BeWX; (C) the amount of β -1 \rightarrow 4-xylobiose released from WAX. Hydrolysis was performed in NaOAc (pH 5.5) at 30°C for up to 16 h. (D) The amount of β -1 \rightarrow 4-xylobiose released from 4-*O*-methyl-D-glucuronidase (MGX) with or without α -glucuronidase (MGX \pm AGU); (E) the amount of β -1 \rightarrow 4-xylobiose released from WAX treated with or without arabinofuranosidase (WAX \pm ABF). All assays were performed in duplicate. The annotation was based on standards and previous research [52].

Table 2 Putative subsite residues of the GH30_7 enzymes based on the comparison of homology models and reported structures^a.

Enzyme name	PDB ID	Ligand	Catalytic site	Putative subsite -1	Putative subsite -2	Short loop (putative subsite -3)
Template						
<i>TcXyn30B</i> (GH30_7)	6KRN	GlcA-X2	E202, E297	W141, N201, E202, Y279, E297, L301	F44, R46, Y209, W341, E345, S347, T349, S351	T91, S92, N93 , L94, M95, N96
<i>EcXyn30A</i> (GH30_8)	2Y24	GlcA-X3	E165, E253	W113, N164, E165, E253	R293, W55, Y295, S258, Y255, W289	-
Models						
<i>TrXbhA</i>	-	GlcA-X2	E198, E290	W137, N197, E198, Y272, E290, L294	F41, R43, Y205, W333, Q337, P341, T343, E345	N86, S87, D88 , F89 , M90, N91
<i>TrXbhA</i>	-	GlcA-X2	E187, E281	W126, N186, E187, Y263, E281, L285	F29, R31, Y194, W324, E328, G330, S332, S334	I76, K77, D78 , F79 , M80, N81
<i>TrExIA</i>	-	GlcA-X2	E188, E278	W127, N187, E188, Y260, E278, L282	F32, R34, Y195, W321, Q325, G327, T329, S331	S76, S77, D78 , H79 , M80, Y81
<i>PgExIA</i>	-	GlcA-X2	E178, E270	W117, N177, E178, Y252, E270, E274	F29, R31, Y185, W315, E319, Q323, S325, S327	-

^a, The amino acids which could affect the substrate specificity are in boldface.

xylobiose from WAX than from BeWX, while the other enzymes showed only trace amounts or no xylobiose release from WAX (**Figure 5B, C**). This agreed with most of the previously characterized enzymes from GH30_7, which also showed much lower activity against WAX than BeWX [3, 4, 38, 41]. This could be due to the high L-arabinosyl substitution of WAX, which might hinder the binding of these enzymes to WAX.

To confirm the effect of L-arabinosyl substitution towards GH30_7 activity, WAX was treated with GH51 α -arabinofuranosidase (ABF) from *A. niger*, which partially hydrolyzes mono- L-arabinosyl substitutions at O-3 and di- L-arabinosyl substitutions at O-2 and O-3 from WAX [53]. WAX contains about 0.38 mg L-arabinose per mg polysaccharide, of which around 30% L-arabinose was released after incubation with GH51 ABF. The removal of L-arabinosyl substitution in WAX improved the release of xylobiose by all enzymes (**Figure 5E**). *TtXbhA* released 3-fold more xylobiose from ABF-treated WAX, and the other GH30_7 enzymes could also release detectable amount of xylobiose. These results indicate that the high degree of L-arabinosyl substitution hinders the accessibility of the enzyme to the xylan backbone.

Differences in the catalytic region of GH30_7 affect their substrate specificities

To investigate the catalytic mechanism of GH30 fungal enzymes, homology modelling analysis of selected candidates was used in this study. Currently, there is no structure available with >35% amino acid sequence similarity to be used as a reliable template for the homology modeling of GH30_3, GH30_11 and GH30_5 enzymes (**Suppl. Table 3**). Hence, only the homology models of the enzymes from GH30_7 were created (**Suppl. Table 3, Suppl. Fig. 3**). *T. cellulolyticus* *TcXyn30B* (**Suppl. Fig. 3A**, PDB ID: 6KRN) [26] was used as a template for homology models of GH30_7 enzymes (**Suppl. Fig. 3B-3E**). The putative subsites -1 to -3 are predicted based on the structure of *TcXyn30B* as well as the *E. chrysanthemi* *EcXyn30A* (**Suppl. Fig. 3F**; PDB ID: 2Y24) [28].

A comparison of the putative catalytic amino acids of GH30_7 enzymes showed that the residues in subsite -1 are highly conserved, while those in subsites -2 and -3 are less conserved (**Table 2**). Instead of the -3 subsite of *PsExlA*, a short loop was observed with *TtXbhA*, *TeXbhA* and *TtExlA* (**Suppl. Fig. 2, Suppl. Fig. 3B-3E**). This loop forms a steric barrier close to the catalytic site to accommodate two D-xylosyl residues, at subsites -1 and -2, explaining the xylobiose release [3, 41]. Within the loop, two amino acids were reported to possibly contribute to xylobiohydrolase activity, i.e., N93 in *TcXyn30B* (corresponding to D88 of *TtXbhA*, D78 of *TeXbhA*, and D78 of *TtExlA*) and W101 in *AaXyn30A* (corresponding to F89 of *TtXbhA*, F79 of *TeXbhA*, and H79 of *TtExlA*) (**Suppl. Fig. 2, Suppl. Fig. 3**) [3, 26].

Conclusions

A combination of phylogenetic analysis and biochemical characterization of enzymes have led to the discovery of a new fungal subfamily in GH30, GH30_11, which displays β -(1 \rightarrow 6)-galactobiohydrolase activity. The combined approach also helped to identify novel enzyme activities and confirmed that different fungal subfamilies harbored enzymes with distinct substrate specificities. However, while the different subfamilies act on different polysaccharides, they all mainly release 'short' non-digestible di- and oligosaccharides, which could be of interest in the food and feed industries. These findings contribute to understanding the fungal GH30 subfamily and facilitate industrial applications of fungal GH30 enzymes.

Author contributions

RPdV conceived and supervised the overall project. RPdV and AD selected the enzyme candidates from GH30 phylogenetic tree and designed the experiment. MAK and DK recommended and provided some substrates for experiment. XL and AD conducted the experiments. XL, DK and AD

analyzed the data. XL and AD wrote the original draft. All authors commented on the manuscript.

Funding

This work was partly supported by the China Scholarship Council (grant no: 201803250066).

References

1. Shrivastava, S., *Industrial Applications of Glycoside Hydrolases*. Springer, 2020.
2. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013*. *Nucleic Acids Research*, 2014. 42(D1): p. D490-D495.
3. Šuchová, K., et al., *A novel GH30 xylobiohydrolase from Acremonium alcalophilum releasing xylobiose from the non-reducing end*. *Enzyme and Microbial Technology*, 2020. 134: p. 109484.
4. Katsimpouras, C., et al., *A novel fungal GH30 xylanase with xylobiohydrolase auxiliary activity*. *Biotechnology for Biofuels*, 2019. 12(1): p. 1-14.
5. St John, F.J., J.M. González, and E. Pozharski, *Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups*. *FEBS Letters*, 2010. 584(21): p. 4435-4441.
6. Crooks, C., N.J. Bechle, and F.J. St John, *A New Subfamily of Glycoside Hydrolase Family 30 with Strict Xylobiohydrolase Function*. *Frontiers in Molecular Biosciences*, 2021: p. 749.
7. Sakurama, H., et al., *β -Glucuronidase from Lactobacillus brevis useful for baicalin hydrolysis belongs to glycoside hydrolase family 30*. *Applied Microbiology and Biotechnology*, 2014. 98(9): p. 4021-4032.
8. Nierman, W.C., et al., *Genomic sequence of the pathogenic and allergenic filamentous fungus Aspergillus fumigatus*. *Nature*, 2005. 438(7071): p. 1151-1156.
9. Montero, M., et al., *BGN16. 3, a novel acidic β -1, 6-glucanase from mycoparasitic fungus Trichoderma harzianum CECT 2413*. *The FEBS Journal*, 2005. 272(13): p. 3441-3448.
10. Oyama, S., et al., *Cloning and expression of an endo-1, 6- β -D-glucanase gene (neg1) from Neurospora crassa*. *Bioscience, Biotechnology, and Biochemistry*, 2002. 66(6): p. 1378-1381.
11. Konno, N. and Y. Sakamoto, *An endo- β -1, 6-glucanase involved in Lentinula edodes fruiting body autolysis*. *Applied Microbiology and Biotechnology*, 2011. 91(5): p. 1365-1373.
12. Oyama, S., et al., *Functional analysis of an endo-1, 6- β -D-glucanase gene (neg-1) from Neurospora crassa*. *Bioscience, Biotechnology, and Biochemistry*, 2006. 70(7): p. 1773-1775.
13. Okemoto, K., et al., *Purification and characterization of an endo- β -(1 \rightarrow 6)-galactanase from Trichoderma viride*. *Carbohydrate Research*, 2003. 338(3): p. 219-230.
14. Takata, R., et al., *Degradation of carbohydrate moieties of arabinogalactan-proteins by glycoside hydrolases from Neurospora crassa*. *Carbohydrate Research*, 2010. 345(17): p. 2516-2522.
15. Meyer, T.S.M., et al., *Biotechnological production of oligosaccharides—applications in the food industry*. *Food Production and Industry*, 2015. 2: p. 25-78.
16. Patel, S. and A. Goyal, *Functional oligosaccharides: production, properties and applications*. *World Journal of Microbiology and Biotechnology*, 2011. 27(5): p. 1119-1128.
17. Swennen, K., C.M. Courtin, and J.A. Delcour, *Non-digestible oligosaccharides with prebiotic properties*. *Critical Reviews in Food Science and Nutrition*, 2006. 46(6): p. 459-471.
18. Lorito, M., et al., *Synergistic interaction between fungal cell wall degrading enzymes and different antifungal compounds enhances inhibition of spore germination*. *Microbiology*, 1994. 140(3): p. 623-629.
19. Kottom, T.J., et al., *Evidence for proinflammatory β -1, 6 glucans in the Pneumocystis carinii cell wall*. *Infection and Immunity*, 2015. 83(7): p. 2816-2826.
20. Polizeli, M., et al., *Xylanases from fungi: properties and industrial applications*. *Applied Microbiology and Biotechnology*, 2005. 67(5): p. 577-591.
21. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server*. *Nucleic Acids Research*, 2014. 42(W1): p. W320-W324.
22. Gouet, P., X. Robert, and E. Courcelle, *ESPrpt/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins*. *Nucleic Acids Research*, 2003. 31(13): p. 3320-3323.
23. Li, X., et al., *Functional validation of two fungal subfamilies in carbohydrate esterase family 1 by bio-*

- chemical characterization of esterases from uncharacterized branches. *Frontiers in Bioengineering and Biotechnology*, 2020: p. 694.
24. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets*. *Molecular Biology and Evolution*, 2016. 33(7): p. 1870-1874.
 25. Söding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction*. *Nucleic Acids Research*, 2005. 33(suppl_2): p. W244-W248.
 26. Nakamichi, Y., et al., *Substrate recognition by a bifunctional GH30-7 xylanase B from Talaromyces cellulolyticus*. *FEBS Open Bio*, 2020. 10(6): p. 1180-1189.
 27. Linares, N.C., et al., *Recombinant production and characterization of six novel GH27 and GH36 α -galactosidases from Penicillium subrubescens and their synergism with a commercial mannanase during the hydrolysis of lignocellulosic biomass*. *Bioresource Technology*, 2020. 295: p. 122258.
 28. Urbániková, L., et al., *Structural basis for substrate recognition by Erwinia chrysanthemi GH30 glucuronoxylanase*. *The FEBS Journal*, 2011. 278(12): p. 2105-2116.
 29. Dilokpimol, A., et al., *Expanding the feruloyl esterase gene family of Aspergillus niger by characterization of a feruloyl esterase, FaeC*. *New Biotechnology*, 2017. 37: p. 200-209.
 30. Nordberg Karlsson, E., et al., *Endo-xylanases as tools for production of substituted xylooligosaccharides with prebiotic properties*. *Applied Microbiology and Biotechnology*, 2018. 102(21): p. 9081-9088.
 31. Mäkelä, M.R., et al., *Characterization of a feruloyl esterase from Aspergillus terreus facilitates the division of fungal enzymes from Carbohydrate Esterase family 1 of the carbohydrate-active enzymes (CAZy) database*. *Microbial Biotechnology*, 2018. 11(5): p. 869-880.
 32. Broxterman, S.E., G. van Erven, and H.A. Schols, *The solubility of primary plant cell wall polysaccharides in LiCl-DMSO*. *Carbohydrate Polymers*, 2018. 200: p. 332-340.
 33. Yoshida, S., et al., *Structural and functional analyses of a glycoside hydrolase family 5 enzyme with an unexpected β -fucosidase activity*. *Biochemistry*, 2011. 50(16): p. 3369-3375.
 34. Ōmura, S., et al., *Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites*. *Proceedings of the National Academy of Sciences*, 2001. 98(21): p. 12215-12220.
 35. Kotake, T., et al., *Molecular cloning and expression in Escherichia coli of a Trichoderma viride endo-beta-(1-6)-galactanase gene*. *Biochemical Journal*, 2004. 377(3): p. 749-755.
 36. Luo, H., et al., *Molecular cloning and characterization of the novel acidic xylanase XYLD from Bispora sp. MEY-1 that is homologous to family 30 glycosyl hydrolases*. *Applied Microbiology and Biotechnology*, 2010. 86(6): p. 1829-1839.
 37. Espinoza, K. and J. Eyzaguirre, *Identification, heterologous expression and characterization of a novel glycoside hydrolase family 30 xylanase from the fungus Penicillium purpurogenum*. *Carbohydrate Research*, 2018. 468: p. 45-50.
 38. Biely, P., et al., *Trichoderma reesei XYN VI—a novel appendage-dependent eukaryotic glucuronoxylan hydrolase*. *The FEBS Journal*, 2014. 281(17): p. 3894-3903.
 39. Tenkanen, M., et al., *Xylanase XYN IV from Trichoderma reesei showing exo- and endo-xylanase activity*. *The FEBS Journal*, 2013. 280(1): p. 285-301.
 40. Nakamichi, Y., et al., *Mode of action of GH30-7 reducing-end xylose-releasing exoxylanase A (Xyn30A) from the filamentous fungus Talaromyces cellulolyticus*. *Applied and Environmental Microbiology*, 2019. 85(13): p. e00552-19.
 41. Nakamichi, Y., et al., *Structural and functional characterization of a bifunctional GH30-7 xylanase B from the filamentous fungus Talaromyces cellulolyticus*. *Journal of Biological Chemistry*, 2019. 294(11): p. 4065-4078.
 42. Sørensen, S.O., et al., *Pectin engineering: modification of potato pectin by in vivo expression of an endo-1,4- β -D-galactanase*. *Proceedings of the National Academy of Sciences*, 2000. 97(13): p. 7639-7644.
 43. Wefers, D., R. Flörchinger, and M. Bunzel, *Detailed structural characterization of arabinans and galactans of 14 apple cultivars before and after cold storage*. *Frontiers in Plant Science*, 2018: p. 1451.
 44. Fujita, K., et al., *Degradative enzymes for type II arabinogalactan side chains in Bifidobacterium longum subsp. longum*. *Applied Microbiology and Biotechnology*, 2019. 103(3): p. 1299-1310.
 45. Loosveld, A., et al., *Structural variation and levels of water-extractable arabinogalactan-peptide in European wheat flours*. *Cereal Chemistry*, 1998. 75(6): p. 815-819.
 46. Sakka, M., et al., *Characterization of Xyn30A and Axl43A of Bacillus licheniformis SVD1 identified by its genomic analysis*. *Enzyme and Microbial Technology*, 2012. 51(4): p. 193-199.

47. Valenzuela, S.V., P. Diaz, and F.J. Pastor, *Modular glucuronoxylan-specific xylanase with a family CBM35 carbohydrate-binding module*. *Applied and Environmental Microbiology*, 2012. 78(11): p. 3923-3931.
48. Padilha, I.Q., et al., *A glucuronoxylan-specific xylanase from a new Paenibacillus favisporus strain isolated from tropical soil of Brazil*. *International Microbiology*, 2014, vol. 17, num. 3, p. 175-184, 2014.
49. Teleman, A., et al., *Characterization of O-acetyl-(4-O-methylglucurono) xylan isolated from birch and beech*. *Carbohydrate Research*, 2002. 337(4): p. 373-377.
50. Hurlbert, J.C. and J.F. Preston III, *Functional characterization of a novel xylanase from a corn strain of Erwinia chrysanthemi*. *Journal of Bacteriology*, 2001. 183(6): p. 2093-2100.
51. Golan, G., et al., *Crystal structures of Geobacillus stearothermophilus α -glucuronidase complexed with its substrate and products: mechanistic implications*. *Journal of Biological Chemistry*, 2004. 279(4): p. 3014-3024.
52. Martínez, P.M., et al., *The two Rasamsonia emersonii α -glucuronidases, ReGH67 and ReGH115, show a different mode-of-action towards glucuronoxylan and glucuronoxyloligosaccharides*. *Biotechnology for Biofuels*, 2016. 9(1): p. 1-10.
53. McCleary, B.V., et al., *Hydrolysis of wheat flour arabinoxylan, acid-debranched wheat flour arabinoxylan and arabino-xylo-oligosaccharides by β -xylanase, α -L-arabinofuranosidase and β -xylosidase*. *Carbohydrate Research*, 2015. 407: p. 79-96.
54. Nakamichi, Y., et al., *GH30-7 endoxylanase C from the filamentous fungus Talaromyces cellulolyticus*. *Applied and Environmental Microbiology*, 2019. 85(22): p. e01442-19.

Supplementary material

Suppl. Table 1 Fungal species included in the phylogenetic analysis. Related to **Figure 1** and **Suppl. Fig. 1**.

Suppl. Table 2 Amount of mono- and di-saccharides released from different substrates by crude GH30 enzymes. *Available upon request from author.*

Suppl. Table 3 Summary of validation parameters of the homology models in this study.

Suppl. Fig. 1 Full phylogenetic relationship among GH30 members from fungi, bacteria, and animals based on their amino acid sequences (From **Figure 1**). *Available upon request from author.*

Suppl. Fig. 2 Amino acid sequence alignment of GH30_7 candidates with characterized GH30_7 and GH30_8 enzymes. *Available upon request from author.*

Suppl. Fig. 3 Structural models representing the catalytic site region of the crystal structure of *Talaromyces cellulolyticus* (TcXyn30B), *Erwinia chrysanthemi* (EcXyn30A) and homology models of GH30_7 enzymes.

Suppl. Table 1 Fungal species included in the phylogenetic analysis. Related to Figure 1 and Suppl. Fig. 1.

Phylum	Clade	Fungi	Homotypic synonym
Basidiomycete	Agaricomycotina	<i>Trametes versicolor</i> v1.0	
Basidiomycete	Agaricomycotina	<i>Pycnoporus cinnabarinus</i> BRFM 137	
Basidiomycete	Agaricomycotina	<i>Dichomitus squalens</i> CBS464.89 v1.0	
Basidiomycete	Agaricomycotina	<i>Dichomitus squalens</i> LYAD-421 SS1 v1.0	
Basidiomycete	Agaricomycotina	<i>Phlebiopsis gigantea</i> v1.0	
Basidiomycete	Agaricomycotina	<i>Phanerochaete chrysosporium</i> RP-78 v2.2	
Basidiomycete	Agaricomycotina	<i>Phanerochaete chrysosporium</i> v2.0	
Basidiomycete	Agaricomycotina	<i>Pleurotus ostreatus</i> PC15 v2.0	
Basidiomycete	Agaricomycotina	<i>Coprinopsis cinerea</i>	
Basidiomycete	Agaricomycotina	<i>Gymnopus luxurians</i> v1.0	
Basidiomycete	Agaricomycotina	<i>Laccaria bicolor</i> v2.0	
Basidiomycete	Agaricomycotina	<i>Postia placenta</i> MAD-698-R-SB12 v1.0	
Basidiomycete	Agaricomycotina	<i>Stereum hirsutum</i> FP-91666 SS1 v1.0	
Basidiomycete	Agaricomycotina	<i>Lentinula edodes</i>	
Ascomycete	Dothideomycetes	<i>Mycosphaerella graminicola</i> v2.0	
Ascomycete	Dothideomycetes	<i>Zymoseptoria pseudotrifici</i> STIR04_2.2.1	
Ascomycete	Dothideomycetes	<i>Stagonospora nodorum</i> SN15 v2.0	
Ascomycete	Sordariomycetes	<i>Fusarium oxysporum</i> f. sp. <i>lycopersici</i> 4287 v2	
Ascomycete	Sordariomycetes	<i>Thermothelomyces thermophila</i> v2.0	<i>Myceliophthora thermophila</i>
Ascomycete	Sordariomycetes	<i>Podospora anserina</i> S mat+	
Ascomycete	Sordariomycetes	<i>Magnaporthe grisea</i> v1.0	
Ascomycete	Sordariomycetes	<i>Trichoderma reesei</i> v2.0	
Ascomycete	Sordariomycetes	<i>Trichoderma reesei</i> RUT C-30	
Ascomycete	Sordariomycetes	<i>Neurospora crassa</i> IFO6068	
Ascomycete	Sordariomycetes	<i>Neurospora crassa</i> OR74A	
Ascomycete	Sordariomycetes	<i>Trichoderma lixii</i>	
Ascomycete	Sordariomycetes	<i>Trichoderma viride</i>	
Ascomycete	Sordariomycetes	<i>Acremonium alcalophilum</i> v2.0	
Ascomycete	Eurotiomycetes	<i>Aspergillus fumigatus</i> Af293 from AspGD	
Ascomycete	Eurotiomycetes	<i>Aspergillus niger</i> NRRL3	
Ascomycete	Eurotiomycetes	<i>Aspergillus japonicus</i> CBS 114.51 v1.0	
Ascomycete	Eurotiomycetes	<i>Penicillium subrubescens</i> FBCC1632 / CBS132785	
Ascomycete	Eurotiomycetes	<i>Penicillium rubens</i> Wisconsin 54-1255	
Ascomycete	Eurotiomycetes	<i>Penicillium purpureogenus</i>	<i>Talaromyces purpureogenus</i>
Ascomycete	Eurotiomycetes	<i>Talaromyces atroroseus</i>	
Ascomycete	Eurotiomycetes	<i>Talaromyces cellulolyticus</i>	
Ascomycete	Eurotiomycetes	<i>Talaromyces emersonii</i>	<i>Rasamsonia emersonii</i>
Ascomycete	Bispora	<i>Bispora</i> sp. MEY-1	

Suppl. Table 3 Summary of validation parameters of the homology models in this study.

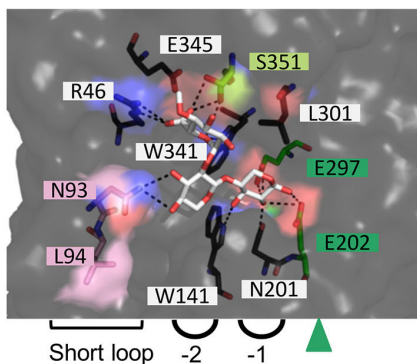
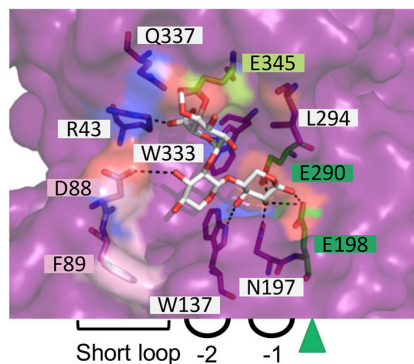
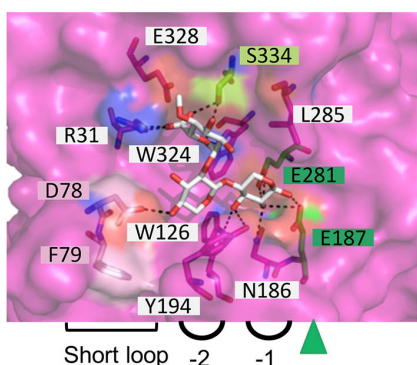
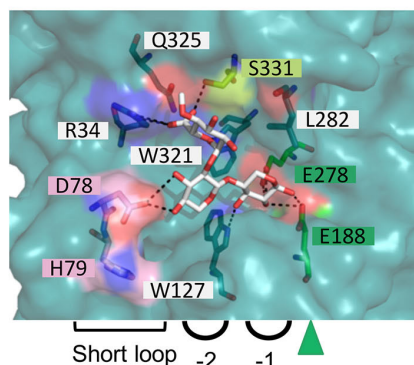
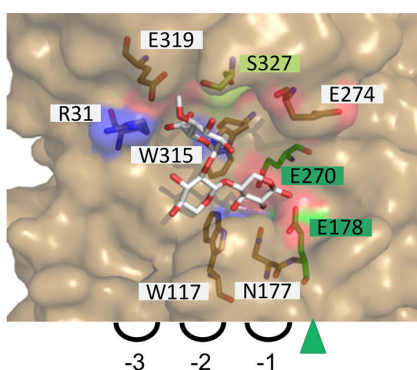
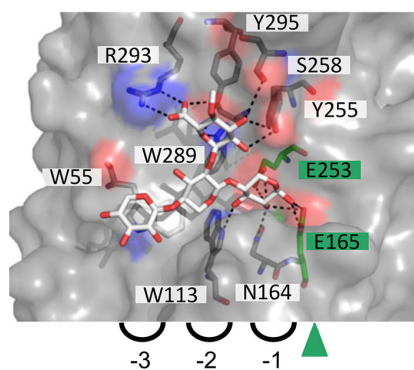
Subfamily	Enzyme name	PDB ID of template	Amino acid sequence identity (%) ^a	RMSD (Å) ^a	Z-score ^b	LGScore / MaxSub ^c	Ramachandran favored regions ^d
	<i>Tx</i> bhA	6KRN_A	46%	1,81	-8.19 (468 aa)	5.697/0.424	89,30%
7	<i>Tz</i> XbhA	6KRN_A	57%	1	-8.57 (459 aa)	5.570/0.415	90,10%
	<i>TE</i> xIA	6KRN_A	45%	1,42	-7.86 (456 aa)	5.675/0.468	88,40%
	<i>Pz</i> ExIA	6KRN_A	39%	2,29	-7.31 (452 aa)	5.288/0.376	88,20%

a, RMSD value that was closer to 0 indicated more closely related to the template structure.

b, Z-scores of > -4.0 indicated good agreement between the model and template structures.

c, LGScore > 4.0 and MaxSub score > 0.3 were considered as good model quality.

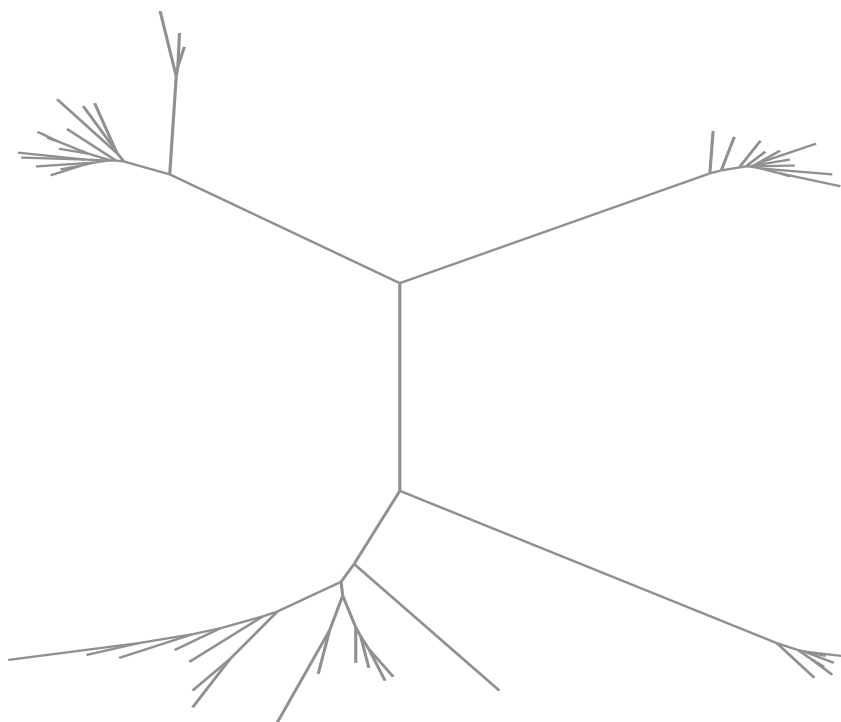
d, Ramachandran favored regions closer to 100% reflect that all amino acids were located inside the favored and energetically allowed regions.

(A) *TcXyn30B* (template) + GlcA-X2(B) *TtXbhA* + GlcA-X2(C) *TeXbhA* + GlcA-X2(D) *TtExlA* + GlcA-X2(E) *PsExlA* + GlcA-X2(F) *EcXyn30A* + GlcA-X3

Suppl. Fig. 3 Structural models representing the catalytic site region of the crystal structure of *Talaromyces cellulosyticus* (*TcXyn30B*), *Erwinia chrysanthemi* (*EcXyn30A*) and homology models of GH30_7 enzymes. Homology models of GH30_7 enzymes were created using *TcXyn30B* as a template (PDB ID: 6KRN) and superimposed with glucuronyl-xylobiose (GlcA-X2, shown as white sticks). The putative subsites (-1 to -3) in the catalytic domain were defined based on *EcXyn30A* complexed with glucuronyl-xylotriose (GlcA-X3, shown as white sticks). The possible hydrogen bonds are indicated by dashed lines. The amino acids which could affect the substrate specificity are highlighted in lime and pink boxes at -2 and -3 subsites, respectively. The catalytic amino acids are highlighted in green boxes and indicated with green triangles.

章节四

系统鉴定青霉菌*P. subrubescens*的九种新型GH51、GH54和GH62 α -L-阿拉伯呋喃糖苷酶



Chapter 4

Comparative characterization of nine novel GH51, GH54 and GH62 α -L-arabinofuranosidases from *Penicillium subrubescens*

This chapter was published in FEBS Letters

Coconi Linares, Nancy, Xinxin Li, Adiphol Dilokpimol, and Ronald P. de Vries

Volume 596, 11 January 2022, P360-368

<https://doi.org/10.1002/1873-3468.14278>

Abstract

α -L-Arabinofuranosidases (ABFs) are important enzymes in plant biomass degradation with a wide range of applications. The ascomycete fungus *Penicillium subrubescens* has more α -L-arabinofuranosidase-encoding genes in its genome compared to other *Penicillia*. We characterized nine ABFs from Glycoside Hydrolase (GH) families GH51, GH54 and GH62 from this fungus and demonstrated that they have highly diverse specificity and activity levels, indicating that the expansion was accompanied by diversification of the enzymes. Comparison of the substrate preference of the enzymes to the expression of the corresponding genes when the fungus was grown on either of two plant biomass substrates did not show a clear correlation, suggesting a more complex regulatory system governing L-arabinose release from plant biomass by *P. subrubescens*.

Introduction

L-Arabinose is a major constituent of plant biomass and is present in the side chains of pectin, xylan and xyloglucan [1]. α -L-Arabinofuranosidases (ABFs) are exo-acting enzymes, which release α -arabinofuranosyl residues from xylan, pectin and xyloglucan, and are therefore an important class of enzymes involved in degradation of plant biomass [1]. They are commonly produced by fungi during growth on plant biomass substrates. ABFs have many applications in plant biomass-based industrial processes, such as increasing digestibility of animal feeds [2], enhancing bread quality and texture [3], clarification of fruit juice [4], increasing aroma in wine [5], promoting pulp delignification [6], improving saccharification for biofuel production [7], and prebiotics production [8].

Based on their amino acid sequence signatures, fungal ABFs have been assigned to four Glycoside Hydrolase (GH) families of the Carbohydrate-Active enZyme database (www.cazy.org) [9]: GH43, GH51, GH54 and GH62. GH43 is a more diverse family that contains various enzyme activities, including enzymes with ABF and/or β -xylosidase activity, while the other families in fungi only contain ABFs. Many studies have reported the differences between ABFs of the different GH families and revealed significant differences also within the specific GH families (see [10] and [11] for a review of these studies).

Genome sequencing has provided an unprecedented insight into the diversity of fungi with respect to their enzymatic potential and revealed significant differences in the number of genes for specific CAZY families [12]. While the evolutionary drivers for this diversity remain to be elucidated, it has been shown that expansion or reduction of enzymes related to a certain polysaccharide correlates with improved or reduced growth of the fungus on that polysaccharide [13-15].

Penicillium subrubescens stands apart from most other *Penicillia*, by an expanded set of genes encoding plant biomass degrading enzymes, in particular with respect to hemicellulase- and pectinase-encoding genes [16]. It was previously shown to be a promising industrial species [17] and has established genome editing methodology [18]. In a previous study, we compared six GH27 and GH36 α -galactosidases, revealing clear differences in substrate specificity and physical properties of the enzymes [19]. Among the expanded gene set of *P. subrubescens* are also genes encoding putative ABFs, five, four and four members of GH51, GH54 and GH62, respectively. While the number of GH51 genes is not unusual, most fungi analyzed so far only have 0-2 GH54 and GH62 genes (unpublished data), suggesting a specific increase in the ability to release L-arabinose from plant biomass by *P. subrubescens*. To determine whether this expansion resulted in an increased set of functional enzymes and whether this led to redundancy or diversification, we compared nine ABFs from *P. subrubescens* in this study by heterologous production and biochemical characterization of the corresponding enzymes. The results demonstrate large variability between the enzymes and demonstrate that the expansion cannot be simply considered to be enzyme redundancy, but appears to have been accompanied with a functional diversification.

Materials and methods

Bioinformatic analysis

Amino acid sequences of all characterized fungal ABFs from GH51, GH54 and GH62 were obtained from the CAZY database (<http://www.cazy.org/>) [9] and combined with the *P. subrubescens* amino acid sequences of candidate secreted ABFs in a multiple sequence alignment using MAFFT v7.0 (<https://www.ebi.ac.uk/Tools/msa/mafft/>). Phylogenetic analysis was computed using the maximum likelihood (ML) method with the Poisson correction distance of substitution rates of the Molecular Evolutionary Genetics Analysis (MEGA v7.0) program [20]. Neighbor joining (NJ) and minimum evolution (ME) trees were conducted both using the Poisson model with uniform rates and complete

deletion. Bootstrap values were generated based on the 500 resampled data sets, using a 50% value as cut-off. All positions containing gaps and missing data were eliminated. The optimal tree from ML method was used as support for the other displayed NJ and ME trees, indicating their bootstrap values in the branches of the ML trees.

Theoretical isoelectric point (pI) and molecular weights (M_w) were calculated by ExPASy–ProtParam tool (https://web.expasy.org/compute_pi/).

cDNA cloning of P. subrubescens ABF encoding genes and production in Pichia pastoris

Specific total RNA was extracted using TRIzol reagent (Invitrogen, Thermo Fischer Scientific, Carlsbad, CA, USA) and purified by NucleoSpin RNA (Macherey-Nagel, Düren, Germany). Full-length cDNA was obtained using ThermoScript Reverse Transcriptase (Invitrogen). Mature ABF encoding cDNAs, without the native signal peptide, were amplified by PCR from RNA obtained from *P. subrubescens* grown on sugar beet pulp as described previously [19]. All PCR products were assembled in pPICZ α A cloning vector using NEBuilder HiFi DNA Assembly Mix (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocol. The resulting plasmids were transformed, propagated into *Escherichia coli* DH5 α competent cells (Invitrogen, Thermo Scientific) and confirmed by sequence analysis (Macrogen, Amsterdam, the Netherlands). After linearization of the plasmids with *Pme*I or *Sac*I (Promega, Madison, WI, USA), the DNA was transformed into *Pichia pastoris* X-33 cells using electroporation.

P. pastoris transformants were selected and cultured, and the proteins subsequently purified as described previously [19]. Fractions containing enzyme were pooled, concentrated, and buffer-exchanged to 20 mM HEPES, pH 7.0, using 10 kDa cut-off ultrafiltration units Amicon (Merck, Darmstadt, Germany). All purification steps were performed at 4°C.

Physical properties of ABFs

The molecular mass of the purified enzymes was estimated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (12% w/v, SDS-PAGE) using Mini-PROTEAN Tetra Cell (Bio-Rad, Hercules, CA, USA) and the standard marker, PageRuler™ Plus Prestained Protein ladder (Thermo Fisher Scientific) with Coomassie Brilliant Blue staining (Bio-Rad). Deglycosylation was performed by treating the native enzymes with endoglycosidase H (New England Biolabs, MA, USA) according to the manufacturer instructions. The protein concentration was determined by a Bradford assay with bovine serum albumin (Pierce, Thermo Scientific, Waltham, MA, USA) as standard.

Enzyme activity assays and enzyme stability

For assessment of α -L-arabinofuranosidase activity, *p*-nitrophenyl- α -L-arabinofuranoside (*p*NP α Ara) (Sigma-Aldrich, Zwijndrecht, The Netherlands) was used as a substrate. The activities were assayed in a total volume of 100 μ L reaction mixtures containing 10 μ L of 2 mM *p*NP α Ara in 50 mM sodium acetate buffer, pH 5.0, and 0.2–0.3 nM purified enzymes at 30°C. The release of *p*-nitrophenol was spectrophotometrically quantified by following the absorbance at 405 nm in a FLUOstar OPTIMA microtiter plate reader (BMG LabTech, Ortenberg, Germany) up to 30 min with a 2 min interval. The linear range was used for calculation of enzyme activity. One unit of enzymatic activity was defined as the amount of protein required to release 1 μ mol of the corresponding product per minute, under the assay condition used.

The effect of pH on the recombinant α -L-arabinofuranosidases was determined over different pH range of 2.0–12.0 using 40 mM Britton-Robinson buffer (adjusted to the required pH) at 30°C, under the conditions described above, except that the reaction was stopped after 30 min with 100 μ L 0.25 M Na₂CO₃. The pH stability was analyzed by incubating the enzymes in the same buffer system in the range from pH 2.0 to pH 12.0 for 1 h and then determining their residual activities

by the standard assay in 50 mM sodium acetate, pH 5.0, at 30°C. The effect of temperature on the recombinant α -L-arabinofuranosidases was determined over the temperature range of 20-90 °C at their optimum pH values, essentially as above. Thermostability was investigated by measuring the enzyme activity remaining after incubation for 1 h at 20-90°C.

Enzyme kinetics

Kinetic parameters of the Michaelis–Menten constant (K_m), maximum enzyme velocity (V_{max}), turnover number (k_{cat}), and the catalytic efficiency (k_{cat}/K_m) were measured by determining the enzyme initial activities over a defined concentration range (0.25-7.0 mM) of *p*NP α Ara. The *p*NP α Ara enzyme initial activities were determined during 30 min using the same experimental and assay conditions described above for each enzyme.

Activity towards plant biomass substrates

Hydrolysis of arabinan was measured using 3 μ g/mL of recombinant enzyme and 1% of wheat arabinoxytan, (sugar beet) arabinan or debranched arabinan (Megazyme), or sugar beet pectin (Pectin Betapec RU301 Herbreith & Fox KG, Neuenbürg, Germany) in 50 mM sodium acetate buffer (pH 4.0). The samples were incubated for 24 h at 30°C and 100 rpm. Saccharification reactions were stopped by incubation at 95°C for 15 min after which the samples were centrifuged (10 min, 4°C, 13 500 g), and the supernatant was diluted 10-fold in milli-Q water prior the analysis. The released L-arabinose was quantified using HPAEC-PAD (Dionex ISC-5000+ system; Thermo Fisher Scientific, Sunnyvale, CA, USA), equipped with a CarboPac PA1 (250 mm \times 4 mm i.d.) column (Thermo Fisher Scientific). The column was pre-equilibrated with 18 mM NaOH followed by a multi-step gradient: 0-20 min: 18 mM NaOH, 20-30 min: 0-40 mM NaOH and 0-400 mM sodium acetate, 30-35 min: 40-100 mM NaOH and 400 mM to 1 M sodium acetate, 35-40 min: 100 mM NaOH and 1 M to 0 M sodium acetate followed by re-equilibration of 18 mM NaOH for 10 min (20°C; flow rate: 0.30 mL/min). 5-250 μ M L-arabinose (Sigma-Aldrich) was used as standards for quantification. The data obtained are the results of two independent biological replicates, and for each replicate, three technical replicates were assayed. The L-arabinose released was calculated as a percentage of the highest hydrolysis reached for each treatment, which was set to 100%.

Results and discussion

Phylogenetic diversity of fungal ABFs

Annotation of the *P. subrubescens* genome predicted five, four and four members of GH51, GH54 and GH62, respectively [16]. However, two of the GH51 members (8514 and 8515) are in fact two parts of a single gene surrounding a gap in the genome sequence, and 4850 (GH54) appears to be an incomplete gene model. These genes were therefore excluded from the study. Naming of the other genes was done in such a way that the genes that were most similar to the well-characterized AbfA, AbfB and AxA from *Aspergillus niger* [21-23] were given the corresponding name, while the other genes were named consecutively.

To evaluate the diversity of these ABFs from *P. subrubescens* in more detail, phylogenetic trees were constructed for each of the families (**Figure 1**) in which all characterized fungal members of the families from the CAZy database (www.cazy.org) were included as well as all the members of a selection of fungal genomes from MycoCosm (<https://myco cosm.jgi.doe.gov/>).

The four GH51 members from *P. subrubescens* separated clearly in the phylogenetic tree (**Figure 1**), with 1-2 members being present in each major branch of ascomycete enzymes. AbfC and AbfH were both present in the same branch, but while AbfH clustered with other Eurotiomycete sequences. AbfC was most similar to an enzyme from the Sordariomycete fungus *Thermothelomyces thermophilus*,

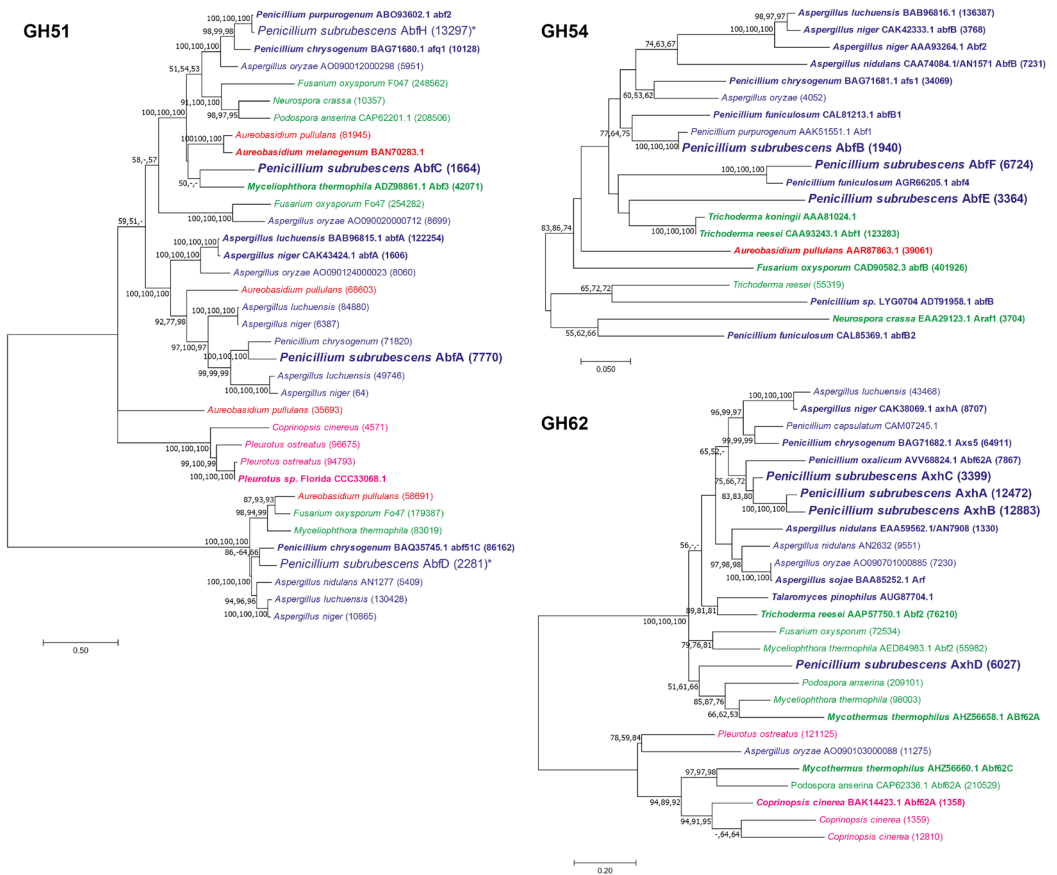


Figure 1 Analysis of phylogenetic relationships among the (putative) fungal α -L-arabinofuranosidases from *Penicillium subrubescens* and selected fungal species (selected based on the presence of characterized enzymes and taxonomic distance to *P. subrubescens*) from GH51, GH54 and GH62. The phylogram was inferred using the Maximum likelihood (ML) method and the optimal tree is shown. Values over 50% bootstrap support (500 replicates) are shown next to the branch nodes from ML (first position), neighbour-joining (NJ, second position) and minimal evolution (ME, third position) tree values from the same data set. All enzymes with biochemical characterization are in boldface. The putative *P. subrubescens* α -arabinofuranosidases that we did not manage to produce and were therefore not biochemically characterized are indicated with an asterisk. The purple highlighted Eurotiomycetes, the green highlighted Sordariomycetes, the red highlighted Dothideomycetes, the pink highlighted Agaricomycetes.

suggesting that *P. subrubescens* may have obtained this gene through horizontal gene transfer. AbfA and AbfD are present in separate branches of the tree, but each within a group of Eurotiomycetes enzymes, indicating that they are evolutionary conserved. The significant distance from each other in the tree indicates a high likelihood for functional differences among the GH51 enzymes of *P. subrubescens*. AbfD appears to be an intracellular enzyme as it lacks a secretory signal sequence. In contrast, the GH54 members from *P. subrubescens* were positioned more closely in their phylogenetic tree, with AbfE and AbfF located in the same branch (**Figure 1**). They are similar to other Eurotiomycete enzymes, although AbfE also had similarity to enzymes from *Trichoderma*. In GH62, three of the four *P. subrubescens* members are clustered very closely which is likely due to recent gene duplications (**Figure 1**). The fourth member (AxDh) is in a branch with enzymes from Sordariomycetes, suggesting that this could also have originated from horizontal gene transfer.

Table 1 Comparison of physical properties and specific activity towards *p*-nitrophenyl- α -L-arabinofuranoside of nine recombinant *Penicillium subrubescens* α -L-arabinofuranosidases. The specific activity was based on the determined molecular mass.

Enzyme Name	Protein Id (JGI)	CAZy Family	pI	Molecular Mass			Specific Activity (U/mg ^a)
				Calculated	Before EndoH	After EndoH	
AbfA	7770	GH51	4.77	67.6	90	68	3.3 ± 0.7
AbfC	1664	GH51	5.18	68.7	90	69	448.5 ± 101.6
AbfB	1940	GH54	5.17	49.4	55	55	507.9 ± 64.1
AbfE	3364	GH54	4.97	50.0	55	55	43.9 ± 7.4
AbfF	6724	GH54	5.84	50.1	55	55	61.2 ± 8.8
AxhA	12472	GH62	4.63	32.8	33	33	4.1 ± 0.5
AxhB	12883	GH62	4.62	33.1	33	33	14.6 ± 2.1
AxhC	3399	GH62	6.39	32.7	33	33	3.8 ± 0.6
AxhD	6027	GH62	5.33	38.8	50	49	10.9 ± 2.8

a, One unit of ABF activity is defined as the amount of protein required to release one μ mol of *p*-nitrophenol per minute.

Functional diversity of *P. subrubescens* ABFs

Using PCR and a cDNA pool of *P. subrubescens* as a template, sequences encoding the mature polypeptide of nine ABFs (AbfA, AbfB, AbfC, AbfE, AbfF, AxhA, AxhB, AxhC, AxhD) were obtained and cloned in-frame with a C-terminal His-tag in *P. pastoris* expression vectors. Despite several attempts, we did not succeed in obtaining a cDNA fragment for AbfH, so this enzyme was not included in the further comparison. The recombinant enzymes were produced in *P. pastoris* and purified (Suppl. Fig. 1), and their physical properties and specific activity towards *p*NP α Ara were compared (Table 1). The molecular mass of the enzymes (after deglycosylation) matched the calculated values. Only for AxhD, a larger molecular mass was observed on gel (Suppl. Fig. 1).

All ABFs had slightly acidic pIs, varying from 4.62 to 6.39 (Table 1) as well as an acidic pH optimum, varying between pH 4 and 5, with the GH62 enzymes having a slightly higher pH optimum (Figure 2A). In contrast, most GH51 and GH54 ABFs have a higher temperature optimum (50°C) than the GH62 ABFs (40°C; Figure 2B). Most enzymes are stable between pH 3 and pH 7, but AxhA and AxhB maintain a high stability at pH 8, while AxhD and AbfF retain more than 50% of their activity at pH 2 (Figure 2C). AbfA and AbfE have the highest temperature stability and maintain 100% of their activity up to 50°C, while the other enzymes maintain this until 40°C (Figure 2D). The similar physical properties are likely related to the natural habitat of *P. subrubescens*, which is a mesophilic species commonly found in soil [24]. Similar values have also been observed for ABFs from other mesophilic species (Suppl. Table 1), such as *Penicillium purperogenum* [25], *Penicillium chrysogenum* [26-28], *A. niger* [21, 29] and *Aureobasidium pullulans* [30].

Differences were observed between the *P. subrubescens* ABFs in their activity on the model substrate *p*NP α Ara, for which AbfB and AbfC had the highest activity, while very low activity was determined for AbfA, AxhA and AxhC (Table 1). Previously it was shown that GH62 ABFs have very low or no activity on this substrate [23], supporting our results. The large differences within GH51 and GH54 are noteworthy as this may provide insight into the efficiency of different enzymes of these families. When compared to other characterized ABFs, the *P. subrubescens* enzymes have similar physical properties.

Although all enzymes have affinities around or below 1 mM, based on their K_m -values, the enzymes display significant differences in their kinetics (Table 2). AbfB has the highest and AbfA the lowest affinity for *p*NP α Ara. AbfB also has the highest catalytic efficiency for this substrate and AxhC the

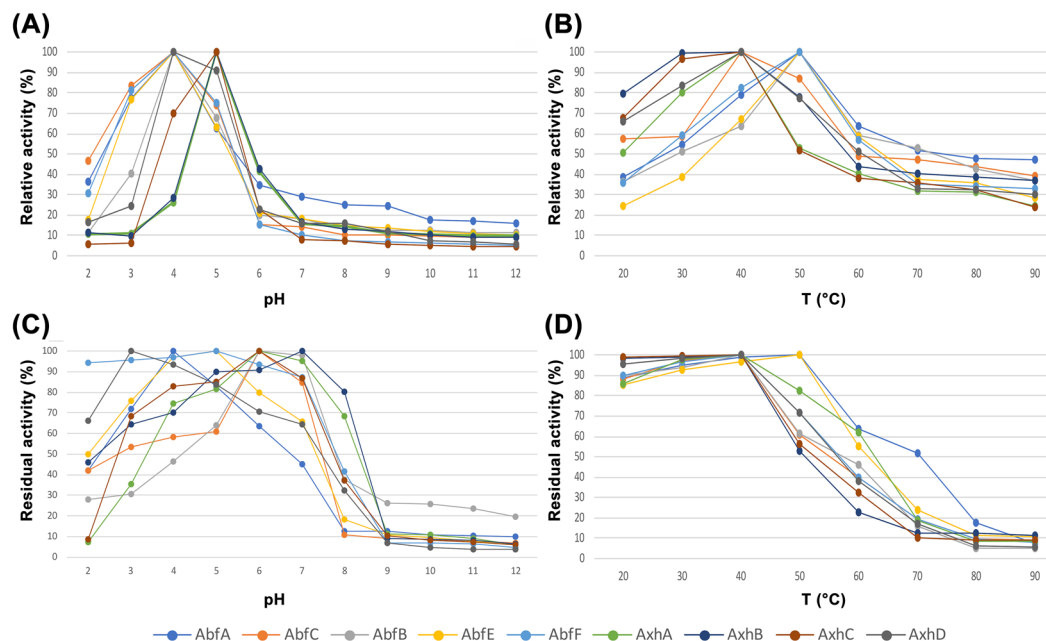


Figure 2 pH and temperature profiles of the recombinant α -L-arabinofuranosidases produced in *Pichia pastoris*. Effect of pH (A) and temperature (B) on the activity and stability of recombinant α -L-arabinofuranosidases using *p*-nitrophenyl- α -L-arabinofuranoside (*p*NP α Ara) as substrate. The pH and temperature-dependence for activity was evaluated at 30°C in 40 mM Britton-Robinson buffer, pH 2.0-12.0, or in 50 mM sodium acetate, pH 5.0, at 20-90°C, respectively. The pH and temperature stability (C, D) was deduced from the residual activity after 1 h incubation. All assays were carried out in triplicate.

lowest. The values for AbfB are in the same range as those observed for its ortholog from *Penicillium purpurogenum* [31]. In general, GH62 ABFs had a lower catalytic efficiency for *p*NP α Ara than most of the GH51 and GH54 ABFs (Table 2), which confirms earlier reports regarding GH62 ABFs. This confirms previous reports that demonstrated that GH62 ABFs belong to type B, which are more active on polymeric arabinoxylan than on unbranched arabinan and *p*NP α Ara [27, 32].

P. subrubescens ABFs display highly diverse activities on natural substrates

The activity of the ABFs was also tested on four polymeric plant biomass substrates that have been previously used to characterize other fungal ABFs (Suppl. Table 1): wheat arabinoxylan, sugar beet pectin, arabinan and debranched arabinan. All enzymes, except AbfA, released the highest amount of L-arabinose from arabinan and this reduced strongly from debranched arabinan (Figure 3). AbfB showed the highest activity on all substrates. Compared to AbfB, AbfA had the lowest activity on all substrates, which matches its low activity on *p*NP α Ara (Table 1). It was unusual in that it had a higher L-arabinose release from debranched arabinan than from (branched) arabinan. AbfC had similar activity to AbfB on arabinan and debranched arabinan, but lower activity on wheat arabinoxylan and sugar beet pectin (Figure 3), indicating a higher specificity for arabinan, which matches previous reports on GH51 and GH54 ABFs [33]. AbfE and AbfF are members of the same branch of the phylogenetic tree and have highly similar activities on the polysaccharide substrates, with a clear preference for pectin-related substrates. The four GH62 enzymes can be divided into two group with respect to their activity on wheat arabinoxylan and sugar beet pectin. While AxhA and AxhC are clearly more active on wheat arabinoxylan, AxhB and AxhD have similar activity on wheat arabinoxylan and sugar beet pectin (Figure 3). Activity on pectin-related arabinan has also been reported for related enzymes, such as Abf62A from *Penicillium oxalicum* [34] and Abf62A and

Table 2 Kinetic parameters for hydrolysis of *p*-nitrophenyl- α -L-arabinofuranoside (*p*NP α Ara) catalyzed by recombinant α -L-arabinofuranosidases from *Penicillium subrubescens*. Parameters were calculated from the initial velocities of *p*-nitrophenol released from *p*NP α Ara at different substrate concentrations.

Enzyme Name	Protein Id (JGI)	CAZy Family	<i>p</i> NP α Ara		
			Km (mM)	Kcat (s ⁻¹)	Kcat/Km (mM ⁻¹ s ⁻¹)
AbfA	7770	GH51	1.469	692.3	471.3
AbfC	1664	GH51	0.099	925.3	9348.1
AbfB	1940	GH54	0.080	1169.1	14599.6
AbfE	3364	GH54	0.681	506.4	744.1
AbfF	6724	GH54	0.620	600.0	968.1
AxhA	12472	GH62	0.930	426.1	458.2
AxhB	12883	GH62	0.662	271.4	410.3
AxhC	3399	GH62	1.088	339.7	312.2
AxhD	6027	GH62	0.720	490.8	681.7

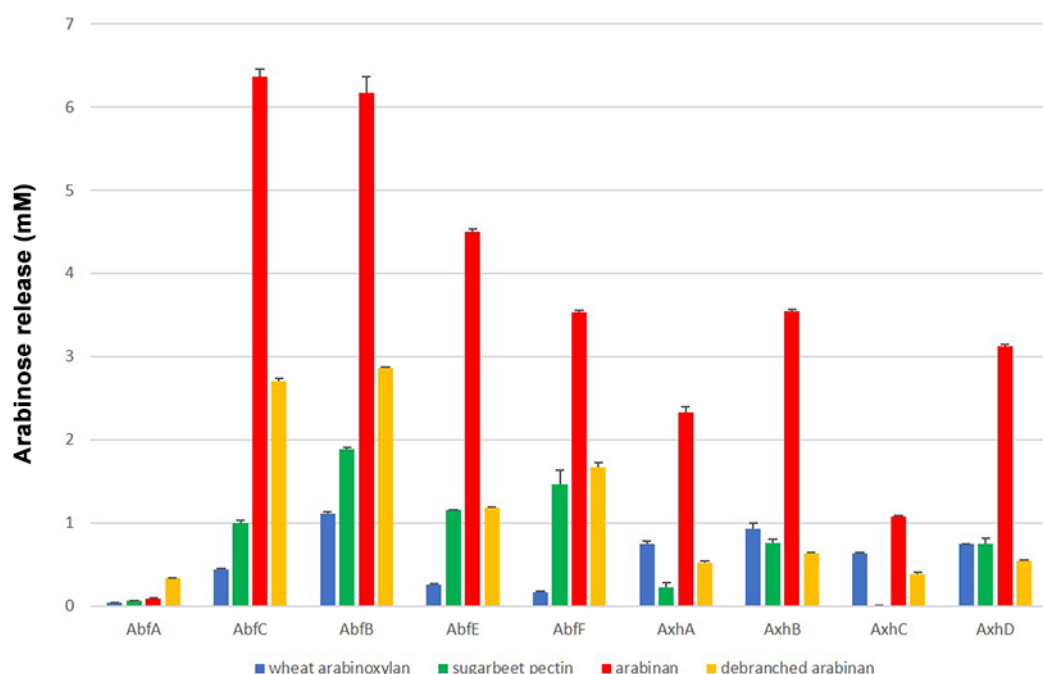


Figure 3 Arabinose-containing lignocellulosic substrate hydrolysis by recombinant α -L-arabinofuranosidases. Blue bar, wheat arabinoxylan; green bar, sugar beet pectin; red bar, arabinan; yellow bar, debranched arabinan. Substrates (1% w/v) were incubated with 3 μ g/mL of recombinant enzyme at 30°C for 24 h. The relative arabinose released was calculated as a percentage of the highest hydrolysis reached for each treatment, which was set to 100%. Values are represented as mean values \pm SD (n = 2).

Abf62C from *Mycothermus thermophilus* [35]. It has been suggested that this activity is particularly related to the α -1,3-L-arabinosyl decorations of the α -1,5-L-arabinosyl backbone or arabinan [32]. However, as all tested GH62 ABFs in our study also hydrolyzed unbranched arabinan, they also have activity on α -1,5-linked L-arabinose.

A previous study described the transcriptome response of *P. subrubescens* to two plant biomass

substrates: wheat bran (rich in arabinoxylan) and sugar beet pulp (rich in pectin that contains arabinan) [36]. We therefore analyzed the expression of the ABF-encoding genes from *P. subrubescens* this data set (**Suppl. Fig. 2**). This revealed that the highest expressed genes were *abfB* and *axhD* (**Suppl. Fig. 2**), while no expression was observed for *abfG*, and very low expression for *axhA* and *axhB*. For most genes, expression was higher on sugar beet pulp than on wheat bran. In contrast, highly similar expression levels on both substrates were observed for *axhD*. This reveals that the substrate profile of the enzymes and expression profile of the corresponding genes do not clearly correlate. AbfB and AbfC are the most active enzymes (**Table 1, Figure 3**), but while *abfB* is the highest expressed gene on sugar beet pulp, *abfC* is only lowly expressed on both biomass substrates (**Suppl. Fig. 2**). The four GH62 ABFs are all very active on wheat arabinoxylan (**Figure 3**), but no preferred expression on wheat bran could be observed (**Suppl. Fig. 2**). Only *axhD* had similar expression levels on both substrates, while *axhC* was higher expressed on sugar beet pulp, and *axhA* and *axhB* had very low expression on both substrates. A possible explanation for this could be that these substrates are still less complex and diverse, and our results therefore do not reflect the natural response of *P. subrubescens*. Regulation of the expression of genes encoding plant biomass degrading enzymes is highly complex in filamentous fungi. Previous studies in *A. niger* [37] and *P. oxalicum* [38] revealed the complex interplay between several regulators, while it has also been shown that orthologous regulators can have significantly different effects in different species [39]. A better understanding of the regulatory network governing plant biomass conversion in *P. subrubescens* will be needed to shed more light on these expression and activity profiles.

Conclusion

In this study, we demonstrated that nine ABF-encoding genes from *P. subrubescens* encode functional enzymes with diverse expression patterns and substrate specificity. In contrast, their physical properties are highly similar, likely driven by the native habitat of *P. subrubescens*. We have now analyzed two expanded enzyme classes (ABFs and AGLs [19]) in this fungus, both of which demonstrate to consist of active proteins with diverse functionalities, suggesting that the expansion in gene numbers is accompanied by a diversification of enzyme function. This indicates that the expansion of these and other enzyme classes in the *P. subrubescens* genome reflects an evolutionary adaptation towards a more diverse and flexible enzymatic toolbox for the degradation of hemicellulose and pectin. This makes this species and others with similar expansions highly interesting candidates for biotechnological applications.

Author contributions

NCL: Investigation, Formal analysis, Writing – review & editing. XL: Investigation, Data curation, Formal analysis, Writing – review & editing. AD: Supervision, Writing – review & editing. RPDV: Conceptualization, Supervision, Funding acquisition, Project administration, Resources, Writing – original draft, review & editing.

Acknowledgements

This study was supported by the National Council of Science and Technology of Mexico (CONACyT) for financial support (Grant No. 263888) to NCL. This work was partly supported by the China Scholarship Council (grant no: 201803250066) to XL.

References

1. de Vries, R.P. and J. Visser, *Aspergillus enzymes involved in degradation of plant cell wall polysaccharides*.

Microbiology and molecular biology reviews, 2001. 65: p. 497-522.

2. Cozannet, P., et al., *Next-generation non-starch polysaccharide-degrading, multi-carbohydrase complex rich in xylanase and arabinofuranosidase to enhance broiler feed digestibility*. **Poultry Science**, 2017. 96(8): p. 2743-2750.
3. Zhou, T., et al., *Improvement of the characteristics of steamed bread by supplementation of recombinant α -L-arabinofuranosidase containing xylan-binding domain*. **Food Biotechnology**, 2019. 33(1): p. 34-53.
4. Churms, S.C., et al., *An L-arabinan from apple-juice concentrates*. **Carbohydrate Research**, 1983. 113(2): p. 339-344.
5. Belda, I., et al., *Microbial contribution to wine aroma and its intended use for wine quality improvement*. **Molecules**, 2017. 22(2).
6. Numan, M.T. and N.B. Bhosle, *α -L-arabinofuranosidases: the potential applications in biotechnology*. **Journal of Industrial Microbiology and Biotechnology**, 2006. 33(4): p. 247-60.
7. Chadha, B.S., A. Monga, and H.S. Oberoi, *α -L-Arabinofuranosidase from an efficient hemicellulolytic fungus *Penicillium janthinellum* capable of hydrolyzing wheat and rye arabinoxylan to arabinose*. **Journal of microbiology, biotechnology and food sciences**, 2017. 6(5): p. 1132-1139.
8. Falck, P., et al., *Arabinoxylanase from glycoside hydrolase family 5 is a selective enzyme for production of specific arabinoxylooligosaccharides*. **Food Chemistry**, 2018. 242: p. 579-584.
9. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013*. **Nucleic Acids Research**, 2014. 42(Database issue): p. D490-495.
10. Lagaert, S., et al., *β -Xylosidases and α -L-arabinofuranosidases: accessory enzymes for arabinoxylan degradation*. **Biotechnology Advances**, 2014. 32(2): p. 316-32.
11. Seiboth, B. and B. Metz, *Fungal arabinan and L-arabinose metabolism*. **Applied Microbiology and Biotechnology**, 2011. 89(6): p. 1665-73.
12. de Vries, R.P. and M.R. Makela, *Genomic and postgenomic diversity of fungal plant biomass degradation approaches*. **Trends in Microbiology**, 2020. 28(6): p. 487-499.
13. Amselem, J., et al., *Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea**. **PLOS Genetics**, 2011. 7(8): p. e1002230.
14. de Vries, R.P., et al., *Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus**. **Genome Biology**, 2017. 18(1): p. 28.
15. Espagne, E., et al., *The genome sequence of the model ascomycete fungus *Podospora anserina**. **Genome Biology**, 2008. 9: p. R77.
16. Peng, M., et al., *The draft genome sequence of the ascomycete fungus *Penicillium subrubescens* reveals a highly enriched content of plant biomass related CAZymes compared to related fungi*. **Journal of Biotechnology**, 2017. 246: p. 1-3.
17. Mäkelä, M.R., et al., **Penicillium subrubescens* is a promising alternative for *Aspergillus niger* in enzymatic plant biomass saccharification*. **Nature Biotechnology**, 2016. 33: p. 834-841.
18. Salazar-Cerezo, S., et al., *CRISPR/Cas9 technology enables the development of the filamentous ascomycete fungus *Penicillium subrubescens* as a new industrial enzyme producer*. **Enzyme and Microbial Technology**, 2020. 133: p. 109463.
19. Coconi Linares, N., et al., *Recombinant production and characterization of six novel GH27 and GH36 α -galactosidases from *Penicillium subrubescens* and their synergism with a commercial mannanase during the hydrolysis of lignocellulosic biomass*. **Bioresource Technology**, 2020. 295: p. 122258.
20. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets*. **Molecular Biology and Evolution**, 2015. 33: p. 1870-1874.
21. Flipphi, M.J.A., et al., *Cloning and characterisation of the *abfB* gene coding for the major α -L-arabinofuranosidase (*AbfB*) of *Aspergillus niger**. **Current Genetics**, 1993. 24: p. 525-532.
22. Flipphi, M.J.A., et al., *Cloning of the *Aspergillus niger* gene encoding α -L-arabinofuranosidase A*. **Applied Microbiology and Biotechnology**, 1993. 39: p. 335-340.
23. Gielkens, M.M.C., J. Visser, and L.H. de Graaff, *Arabinoxylan degradation by fungi: Characterisation of the arabinoxylan arabinofuranohydrolase encoding genes from *Aspergillus niger* and *Aspergillus tubingenensis**. **Current Genetics**, 1997. 31: p. 22-29.
24. Mansouri, S., et al., **Penicillium subrubescens*, a new species efficiently producing inulinase*. **Antonie Van Leeuwenhoek**, 2013. 103(6): p. 1343-1357.
25. Fritz, M., et al., *A family 51 α -L-arabinofuranosidase from *Penicillium purpurogenum*: purification, properties and amino acid sequence*. **Mycological Research**, 2008. 112(Pt 8): p. 933-42.

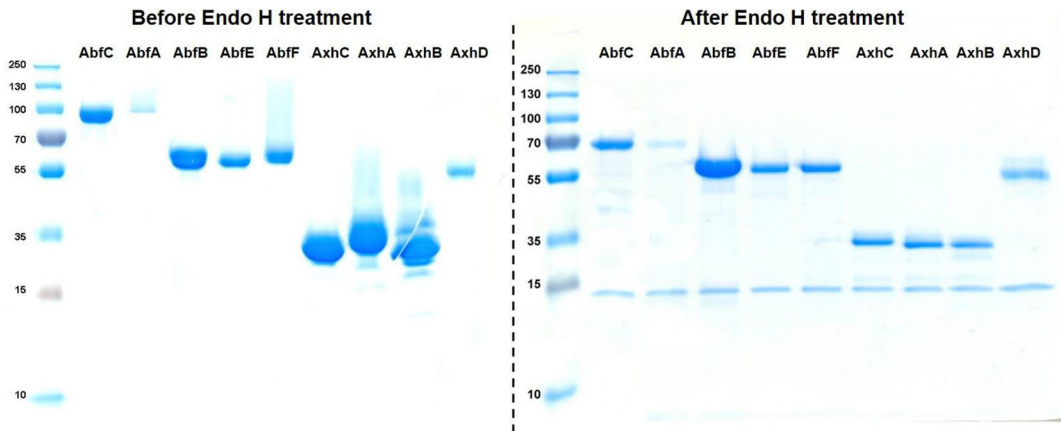
26. Sakamoto, T. and H. Kawasaki, Purification and properties of two type-B alpha-L-arabinofuranosidases produced by *Penicillium chrysogenum*. *Biochimica et Biophysica Acta*, 2003. 1621(2): p. 204-10.
27. Sakamoto, T., et al., Identification of a GH62 alpha-L-arabinofuranosidase specific for arabinoxylan produced by *Penicillium chrysogenum*. *Applied Microbiology and Biotechnology*, 2011. 90(1): p. 137-46.
28. Shinozaki, A., et al., Identification and characterization of three *Penicillium chrysogenum* alpha-L-arabinofuranosidases (PcABF43B, PcABF51C, and AFQ1) with different specificities toward arabinoo-oligosaccharides. *Enzyme and Microbial Technology*, 2015. 73-74: p. 65-71.
29. Alias, N.I., et al., Expression optimisation of recombinant alpha-L-arabinofuranosidase from *Aspergillus niger* ATCC 120120 in *Pichia pastoris* and its biochemical characterisation. *African Journal of Biotechnology*, 2011. 10(35): p. 6700-6710.
30. de Wet, B.J., et al., Characterization of a family 54 alpha-L-arabinofuranosidase from *Aureobasidium pululans*. *Applied Microbiology and Biotechnology*, 2008. 77(5): p. 975-83.
31. Ravanal, M.C. and J. Eyzaguirre, Heterologous expression and characterization of alpha-L-arabinofuranosidase 4 from *Penicillium purpurogenum* and comparison with the other isoenzymes produced by the fungus. *Fungal Biology*, 2015. 119(7): p. 641-7.
32. Wilkens, C., et al., GH62 arabinofuranosidases: Structure, function and applications. *Biotechnology Advances*, 2017. 35(6): p. 792-804.
33. Sakamoto, T. and H. Kawasaki, Purification and properties of two type-B alpha-L-arabinofuranosidases produced by *Penicillium chrysogenum*. *Biochimica et Biophysica Acta*, 2003. 1621(2): p. 204-10.
34. Hu, Y., et al., Cloning and expression of a novel alpha-L,3-arabinofuranosidase from *Penicillium oxalicum* sp. 68. *AMB Express*, 2018. 8(1): p. 51.
35. Kaur, A.P., et al., Functional and structural diversity in GH62 alpha-L-arabinofuranosidases from the thermophilic fungus *Scytalidium thermophilum*. *Microbial Biotechnology*, 2015. 8(3): p. 419-33.
36. Dilokpimol, A., et al., *Penicillium subrubescens* adapts its enzyme production to the composition of plant biomass. *Bioresource Technology*, 2020. 311: p. 123477.
37. Kun, R.S., et al., Blocking utilization of major plant biomass polysaccharides leads *Aspergillus niger* towards utilization of minor components. *Microbial Biotechnology*, 2021. 14(4): p. 1683-1698.
38. Yao, G., et al., Redesigning the regulatory pathway to enhance cellulase production in *Penicillium oxalicum*. *Biotechnology for Biofuels*, 2015. 8: p. 71.
39. Klaubauf, S., et al., Similar is not the same: Differences in the function of the (hemi-)cellulolytic regulator *XlnR (Xlr1/Xyr1)* in filamentous fungi. *Fungal Genetics and Biology*, 2014. 72: p. 73-81.

Supplementary material

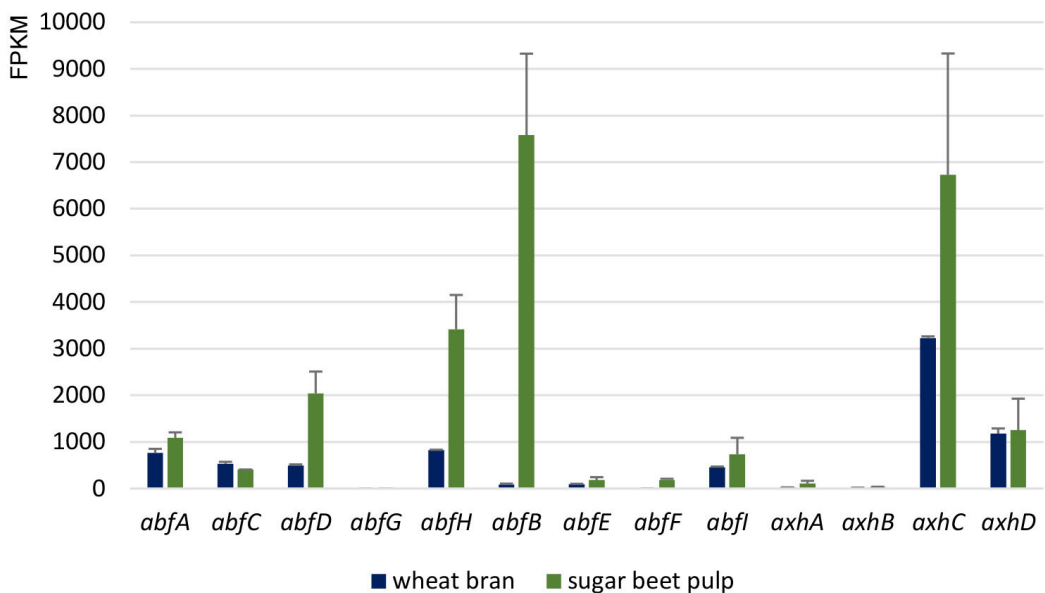
Suppl. Table 1 Comparison of the physical properties of fungal α -L-arabinofuranosidases. The enzymes are organized based on their position in the phylogenetic tree (Fig.1). **Available upon request from author.**

Suppl. Fig. 1 SDS/PAGE analysis of the purified α -L-arabinofuranosidases from *Penicillium subrubescens* before and after deglycosylation with Endo H.

Suppl. Fig. 2 Expression of *Penicillium subrubescens* α -L-arabinofuranosidase-encoding genes on wheat bran and sugar beet pulp.



Suppl. Fig. 1 SDS-PAGE analysis of the purified α -L-arabinofuranosidases from *Penicillium subrubescens* before and after deglycosylation with Endo H.

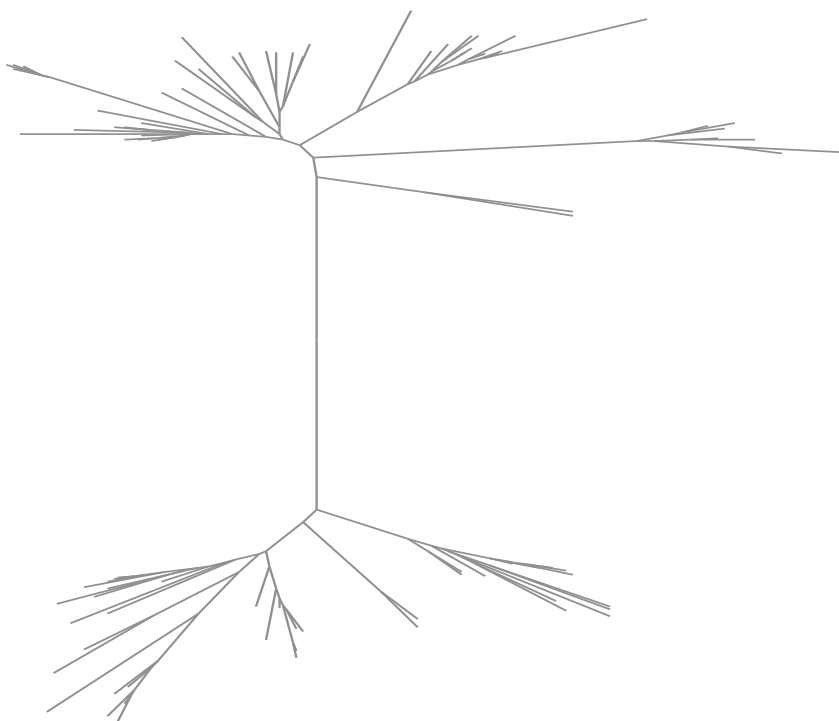


Suppl. Fig. 2 Expression of *Penicillium subrubescens* α -L-arabinofuranosidase-encoding genes on wheat bran and sugar beet pulp [1].

[1] Dilokpimol, A. et al. (2020). *Penicillium subrubescens* adapts its enzyme production to the composition of plant biomass. *Bioresour Technol* 311, 123477.

章节五

青霉菌 *P. subrubescens* GH10和GH11 家族的木聚糖酶：系统比较及其与相同来源的GH51、GH54、GH62 α -L-阿拉伯呋喃糖苷酶的协同作用



Chapter 5

GH10 and GH11 endoxylanases in *Penicillium subrubescens*: comparative characterization and synergy with GH51, GH54, GH62 α -L-arabinofuranosidases from the same fungus

This chapter was published in New Biotechnology

Li, Xinxin, Dimitrios Kouzounis, Mirjam A. Kabel, Ronald P. de Vries

<https://doi.org/10.1016/j.nbt.2022.05.004>

Abstract

Penicillium subrubescens has an expanded set of genes encoding putative endoxylanases (*Ps*XLNs) compared to most other *Penicillia* and other fungi. In this study, all GH10 and GH11 *Ps*XLNs were produced heterologously in *Pichia pastoris* and characterized. They were active towards beech wood xylan (BeWX) and wheat flour arabinoxylan (WAX), and showed stability over a wide pH range. Additionally, *Ps*XLNs released distinct oligosaccharides from WAX, and showed significant cooperative action with *P. subrubescens* α -L-arabinofuranosidases (*Ps*ABFs) from GH51 or GH54 for WAX degradation, giving insight into a more diverse XLN and ABF system for the efficient degradation of complex hemicelluloses. Homology modelling analysis pointed out differences in the catalytic center of *Ps*XLNs, which are discussed in view of the different modes of action observed. These findings facilitate understanding of structural requirements for substrate recognition to contribute to recombinant XLN engineering for biotechnological applications.

Introduction

Xylan is the most abundant hemicellulose present in many types of lignocellulosic plant biomass, forming a complex matrix with lignin within a cellulose fibril [1]. This network is essential for the structural integrity of plants and provides resistance to pathogenic attack, pests and from enzymatic degradation [2]. Degrading xylan to overcome lignocellulose recalcitrance is essential for the efficient utilization of lignocellulose in industry. Endoxylanases (XLNs) are the most crucial enzymes for cleavage of the xylan backbone resulting in a release of xylooligosaccharides (XOS) [3]. Based on amino acid sequence similarity, fungal XLNs have been mainly classified as members of the Glycoside Hydrolase (GH) families GH10 and GH11 in the Carbohydrate-Active enZyme (CAZy) database (<http://www.cazy.org/>) [4]. They differ in their structure and substrate specificity [5-8].

In general, GH10 XLNs have a molecular mass of over 40 kDa and display a salad bowl-like (β/α)₈ barrel catalytic domain, with a shallow groove active site located at the larger radius on the bowl top [7, 9, 10]. GH11 XLNs are less than 30 kDa in size and show a hand-like β -jelly roll structure, with a palm active site situated between the fingers and the thumb [7, 11]. In the catalytic domain, GH10 XLNs have between four and seven substrate-binding subsites, with L-arabinosyl substitutions primarily accommodated at the -3, -2 and +1 subsites and sometimes at the +2 subsite [9, 10]. GH11 XLNs mostly contain five or six subsites, with the -3, +2 and +3 subsites having the ability to tolerate L-arabinosyl substitution [7, 11]. The different tolerance of GH10 and GH11 XLNs to L-arabinosyl substitution explains their distinct substrate specificities and product profiles. GH10 XLNs can tolerate a higher degree of substitution on the xylan chain than GH11 XLNs, and typically result in a greater yield of shorter oligosaccharide products [6, 12]. However, both GH10 and GH11 XLNs have applications in various industries, e.g., food and feed, biofuel production, pulp and paper, and medical and pharmacological industries [13-16].

Regardless of these known differences for GH10 and GH11 XLNs, it is intriguing that some filamentous fungi possess an expanded set of XLNs of both families encoded in their genome. One of such fungi with an expanded set of XLNs is *Penicillium subrubescens* FBCC1632/CBS132785. This fungus has 10 XLN-encoding genes, a number significantly higher than that found in related fungi, e.g., *Aspergillus niger* (gene number: 5), *Trichoderma reesei* (4), *P. chrysogenum Wisconsin* (4), and *P. oxalicum* (8) [16-19]. Previously in *P. subrubescens*, it was shown that the expansion of the set of α -galactosidases [20] and α -L-arabinofuranosidases (ABFs) [21] resulted in functional diversity. While those studies both addressed exo-acting enzymes, in this study the aim was to reveal whether such functional diversification also occurred for the expanded set of XLNs, or whether this largely resulted in enzymatic redundancy.

For this, the genes encoding three GH10 and seven GH11 XLNs in *P. subrubescens* were heterogeneously expressed in *Pichia pastoris*. The recombinant proteins were characterized, and the product profiles towards wheat flour arabinoxylan (WAX) hydrolysis, followed by homology modelling analysis, were used to understand enzyme interactions with decorated heteroxylan. Moreover, the cooperative action of *P. subrubescens* XLNs (*Ps*XLNs) with *P. subrubescens* ABFs (*Ps*ABF) was assessed to find leads to improve existing enzyme cocktails degrading dietary fiber rich in WAX.

Materials and methods

Phylogenetic analysis of fungal GH10 and GH11 XLNs

All fungal amino acid sequences in this study were obtained from JGI Mycocosm (<https://mycocosm.jgi.doe.gov/mycocosm/home>) [22] and the CAZy database (<http://www.cazy.org/>) [4] (**Suppl. Table 1**). Their signal peptides were predicted using SignalP v5.0 (<http://www.cbs.dtu.dk/services/>

SignalP/) and removed manually [23]. Sequences without signal peptides were aligned with the MAFFT server (<https://www.ebi.ac.uk/Tools/msa/mafft/>) [24] and further analyzed in the Molecular Evolutionary Genetic Analysis software version 7 (MEGA7) [25]. Maximum likelihood (ML) was run in MEGA, using 500 bootstrap re-samplings with 95% partial deletion of gap under the Poisson correction distance of substitution rates. Three characterized plant GH10 XLNs were used as an outgroup of the GH10 tree [26-28], whereas three characterized bacterial GH11 XLNs were used as an outgroup of the GH11 tree [29-31].

Cloning of *P. subrubescens* XLN-encoding genes and transformation to *Pichia pastoris*

Total RNA of *P. subrubescens* was extracted and purified as described previously [20]. Full-length cDNA was obtained with ThermoScript Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) using total RNA as a template. The mature XLN-encoding genes without the native signal peptide-coding sequence were amplified from full-length cDNA using the specific primers (**Suppl. Table 2**). The PCR products were purified and digested with the appropriate restriction enzymes (Promega, Madison, WI, USA). The digested products were ligated into pPICZαA vector and transformed into *Escherichia coli* DH5α competent cells for propagation and sequencing. The recombinant plasmids were subsequently extracted from strains containing DNA of the correct sequence, linearized with *PmeI* (Promega), and then transformed by electroporation using a BioRad GenePulser (BioRad, Hercules, CA, USA) into *P. pastoris* strain X-33.

Production and purification of recombinant XLNs

The *P. pastoris* transformants with the highest production level were selected using colony Western Blot as described previously [32], and further cultured in YPD (yeast peptone dextrose medium) and BMGY (buffered glycerol complex medium) sequentially. Induction was done in BMMY (buffered methanol complex medium) medium at 22°C, 280 rpm with methanol being supplemented to 1% (v/v) every 24 h for 96 h. Culture supernatants were harvested (8000 g, 4°C, 20 min), concentrated by Vivaflow 200 membrane of 10 kDa molecular weight cutoff (Sartorius AG, Goettingen, Germany), and purified using an ÄKTA FPLC device (GE Life Sciences, Uppsala, Sweden) at 4°C. Crude enzymes were loaded onto a HisTrap FF 1 mL column (Cytiva, Marlborough, MA, USA) equilibrated with 20 mM HEPES, 0.4 M NaCl, 20 mM imidazole, pH 7.5, and eluted using a linear gradient of 22-400 mM imidazole in buffer mentioned above at a flow rate of 1.0 mL/min [20]. Fractions containing enzyme were collected, concentrated and buffer-exchanged to 20 mM HEPES (pH 7.0) using 10 kDa cut-off ultrafiltration units Amicon (Merck Millipore, Bedford, MA, USA) [20]. The concentration of purified enzymes were determined using the Bradford method with bovine serum albumin (Pierce, Thermo Fisher Scientific, Loughborough, UK) as a standard (5-1000 µg/mL). These purified enzymes were filtered through 0.22 µm filters and stored at 4°C prior to further analysis biochemical properties and hydrolysis patterns.

XLN activity assays and the effects of temperature and pH

XLN activity was determined by using the DNS (3,5-dinitrosalicylic acid) method [33]. Xylan from beech wood (BeWX, Carl Roth GmbH, Karlsruhe, Germany) and wheat flour (medium) (WAX, Megazyme, P-WAXYM, Wicklow, Ireland) were used as substrates. The standard reaction was performed with 20 µL of diluted purified enzyme and 60 µL of 0.5% (w/v) xylan in 100 mM NaOAc buffer (pH 5.0) at 40°C, 110 rpm, for 10 min. The reaction was then stopped by adding 120 µL DNS reagent, and incubated at 95°C for 10 min. After cooling to room temperature, the absorbance at 540 nm was measured in a 96 well microplate. The reaction systems without enzymes were treated as the control. All experiments were carried out in triplicate. One unit (U) of XLN activity was defined as the amount of enzyme releasing 1 µmol of D-xylose per min under the standard assay condition.

Temperature and pH optima of XLNs were determined at 30–90°C in 100 mM NaOAc buffer, pH 5.0, or at 40°C in 100 mM citric acid-Na₂HPO₄ buffer, pH 3.0–8.0, respectively. Thermal and

pH stability of XLNs were estimated by measuring the residual enzyme activities under standard condition after 1 h pre-incubation without substrate. For assay of thermal stability, the 1 h pre-incubation was performed at 30–90°C in 100 mM NaOAc buffer (pH 5.0). For assay of pH stability, the 1 h pre-treatment was carried out at 30°C in 100 mM citric acid- Na_2HPO_4 buffer (pH 2.0–8.0), Tris-NaCl buffer (pH 8.0–9.0), and glycine-NaOH buffer (pH 9.0–11.0).

Hydrolysis of WAX by each independent XLN and analysis of product profile

To investigate the product profile of WAX hydrolysis by each individual *Ps*XLN, 5 $\mu\text{g}/\text{mL}$ XLN was incubated with 0.5% (w/v) WAX in 50 mM NaOAc buffer (pH 5.0) at 40°C, 16 h, under continuous agitation. The reaction without the addition of enzymes was treated as control. All reactions were ended by incubating at 99°C for 15 min, and centrifuged for 10 min (20,000 g, 15°C). The supernatants were diluted twice and 20 times prior to their HPSEC-RI (High Performance Size-Exclusion Chromatography with Refractive Index detection) and HPAEC-PAD (High-Performance Anion Exchange Chromatography with Pulsed Amperometric Detection) analysis, respectively.

Molecular weight distribution determination by HPSEC-RI

The molecular weight distribution of XLN-treated and untreated WAX was determined by HPSEC-RI, according to [34]. Analysis was performed with an Ultimate 3000 HPLC System (Dionex Corp., Sunnyvale, CA, USA) equipped 4000AW, 3000AW, and 2500AW TSK-Gel Super columns (6 mm ID \times 150 mm per column, 6 μm), and a TSK Super AW-L guard column (4.6 mm ID \times 35 mm, 7 μm) (Tosoh Bioscience, Tokyo, Japan). The HPLC system was coupled to an ERC Refractomax 520 detector (Biotech AB, Onsala, Sweden). The system was calibrated using a pullulan series of known MW (Sigma Aldrich, St. Louis, MO, USA).

Product profiling by HPAEC-PAD

Mono- and oligosaccharides were profiled by HPAEC-PAD. Analysis was performed with an ICS7000 HPLC system (Dionex), coupled with an ISC7000 ED PAD detector (Dionex) and equipped with a CarboPac™ PA1 IC column (250 mm \times 2 mm i.d.) and a CarboPac™ PA guard column (50 mm \times 2 mm i.d.). 0.1 M sodium hydroxide (NaOH) (A) and 1 M sodium acetate in 0.1 M NaOH (B) were used as mobile phases. Analysis was performed by injecting 10 μL , at 0.3 mL/min (20°C) using the following elution profile: 0–32 min from 0% to 38% B (linear gradient), 32–37 min from 32% to 100% B, 37–42 min at 100% B (isocratic), 42–42.1 min to 100% A (linear gradient) and 42.1–55 min 100% A (isocratic).

For comparison and annotation of peaks, product profiles of WAX digested by commercially available and well characterized *Tm*GH10 (*Thermotoga maritima*, E-XYLATM) and *Np*GH11 (*Neocallimastix patriciarum*, E-XYLNP, [35]), as well as standards (10–20 $\mu\text{g}/\text{mL}$, XOS, A²XX, XA³XX, XA²XX, A²⁺³XX, Megazyme) are included in this study. Additional peaks were putatively annotated based on the HPAEC elution pattern of AXOS, previously reported by [36].

Synergy with ABF for soluble sugar production

To investigate the cooperative action between each individual *Ps*XLN and different *Ps*ABFs, 0.6 μM of each *Ps*XLN was simultaneously incubated with 6 μM *Ps*ABF for the hydrolysis of 0.5% (w/v) WAX in 100 mM NaOAc buffer (pH 5.0) at 40°C, 110 rpm, 24 h. The reaction systems without enzymes or with only a single *Ps*XLN were treated as controls. The amount of reducing sugars released was determined by the DNS method as mentioned above. All the above hydrolysis assays were carried out in triplicate.

Homology modeling and structure analysis

Multiple sequence alignments of putative *Ps*XLNs with characterized XLNs were visualized using the Easy Sequencing in Postscript (<http://escript.ibcp.fr/EScript/EScript/>) [37]. Homology models were generated and evaluated according to the methods described elsewhere [38, 39]. To obtain the enzyme-ligand complex and further explore how enzymes bind to WAX, models were superpositioned with known experimental 3D structures determined with ligands and visualized by PyMOL 2.3.5 (Schrödinger, Inc., New York, NY, USA) [39]. The models of GH10 XLNs were superpositioned with XA³X (subsite -1 to -3) from *Streptomyces olivaceoviridis* E-86 SoXyn10A (PDB ID: 1V6V) and X₄ (subsite +1 to +4) from *Caldicellulosiruptor bescii* CbXyn10C (PDB ID: 5OFK) [9, 40]. The models of GH11 XLNs were superpositioned with XA³XX (subsite -1 to -4) from *S. olivaceoviridis* E-86 SoXyn11A (PDB ID: 7DFN) and X₃ (subsite +1 to +3) from *Trichoderma reesei* TrXyn11A (PDB ID: 4HK8) [11, 41].

Results and discussion

P. subrubescens harbors a high diversity of XLNs from GH10 and GH11

P. subrubescens stands out from most other *Penicillia*, as it contains an expanded set of genes encoding plant biomass-degrading enzymes, including genes encoding putative XLNs [14, 19, 42]. The JGI annotation for fungal genomes suggested that *P. subrubescens* contains three and seven

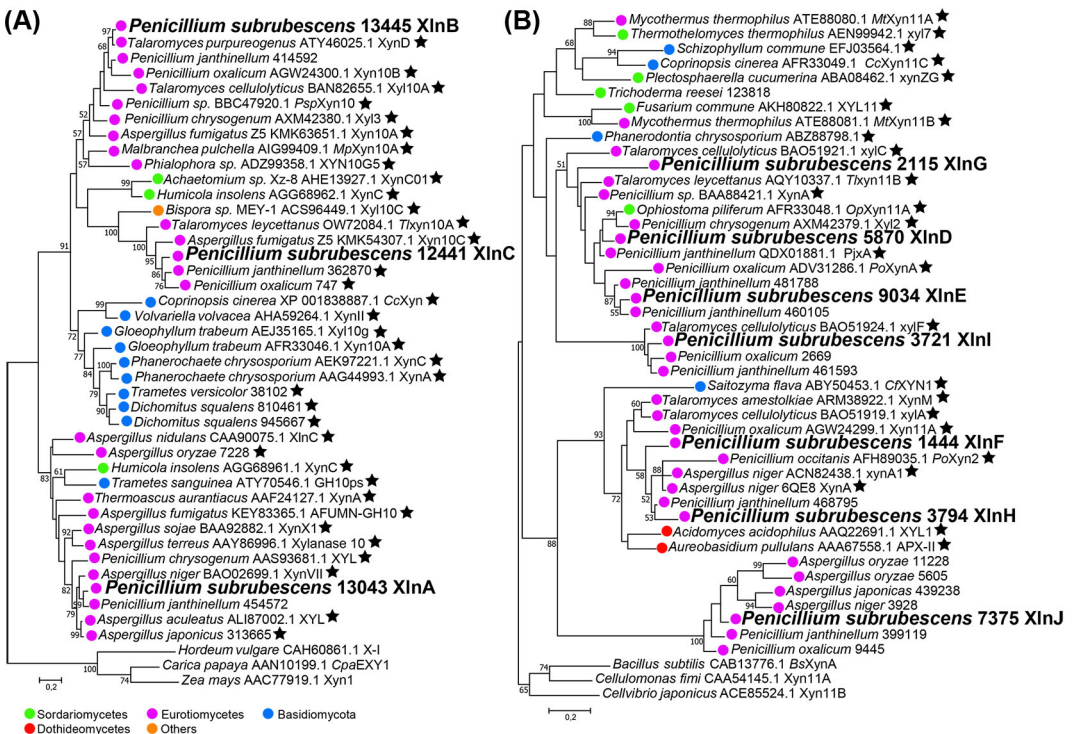


Figure 1 Analysis of phylogenetic relationships among the (putative) fungal XLNs from *Penicillium subrubescens* and selected fungal species from GH10 (A) and GH11 (B). The phylogenetic tree shown was based on the ML method. Only bootstrap values above 50% are shown next to the branches to support nodes. The putative *P. subrubescens* XLNs are highlighted in bold. The previously characterized XLNs are indicated by black stars.

members from GH10 and GH11, respectively, of which one of GH10 and two of GH11 contain a carbohydrate-binding module (CBM) domain in their sequence. To assess the diversity of these putative XLNs in *P. subrubescens* and their evolutionary relationship, *Ps*XLNs from GH10 and GH11 were subjected to phylogenetic analysis with other fungal sequences from the corresponding family, including characterized enzymes (**Figure 1**).

The three GH10 members in *P. subrubescens* showed clear evolutionary distances from each other in the phylogenetic tree, but each was located within a group of Eurotiomycete proteins, particularly from *Penicillia*, indicating that they are evolutionary conserved (**Figure 1A**). *Ps*XlnB with a C-terminal CBM domain clustered together with other Eurotiomycete sequences that contain a CBM domain, and showed very close relationship with *Talaromyces purpureogenus* XynD [43]. The close evolutionary distance between those enzymes may reflect similar properties. The presence of a CBM in XynD was speculated to be possibly involved in the binding of this enzyme to natural lignocellulosic substrates [43]. A similar consideration may be applied to the role of the CBM in *Ps*XlnB.

Similar to GH10, GH11 members in *P. subrubescens* clearly separated in the phylogenetic tree, indicating their phylogenetic diversity, with 3-4 members being present in each main branch of Eurotiomycete sequences (**Figure 1B**). *Ps*XlnG, *Ps*XlnD, and *Ps*XlnE were closer to each other than to any of the other enzymes. *Ps*XlnE possesses a C-terminal CBM and clustered with other sequences from *Penicillium* species. Similarly, *Ps*XlnF and *Ps*XlnH were positioned more closely to each other than to other *Ps*XLNs, and shared a same node with a small number of proteins from Dothideomycetes and Basidiomycetes in addition to those from Eurotiomycetes. The sequence feature of *Ps*XlnI, similar to *Ps*XlnE, harbored a CBM domain. But *Ps*XlnI was located in a separate branch, which clustered very closely with a characterized *T. cellulolyticus* enzyme and with uncharacterized *Penicillium* proteins [44]. *Ps*XlnJ was related to *Ps*XlnF and *Ps*XlnH, which split from the same node. Interestingly, the proteins clustered with *Ps*XlnJ were mainly uncharacterized proteins from *Aspergillus* and other *Penicillium* species.

*Ps*XLNs have diverse biochemical properties

The cDNA fragments encoding XLNs in *P. subrubescens* were overexpressed in *P. pastoris*. Recombinant proteins were produced and purified (**Suppl. Fig. 1**). The functional characterization showed pH optima for *Ps*XLNs ranged from acidic to neutral (pH 4–7), with most *Ps*XLNs having temperature optima of 40°C (**Table 1**). They are similar to the properties of characterized enzymes clustered with *Ps*XLNs in the phylogenetic tree [45-50] (**Figure 1A**). Surprisingly, four *Ps*XLNs (*Ps*XlnB and *Ps*XlnC of GH10, and *Ps*XlnI and *Ps*XlnF of GH11) displayed the highest activity at 70°C (**Table 1**).

All *Ps*XLNs remained stable through the acidic to basic pH range, of which *Ps*XlnB from GH10, and *Ps*XlnH and *Ps*XlnE from GH11 exhibited a significantly wider pH stability than others, as these enzymes retained >60% of their residual activity after 1 h of incubation at pH 11.0 (**Table 1**). This high alkaline tolerance makes *Ps*XLNs attractive for various higher pH biotechnological applications [51, 52]. GH10 *Ps*XLNs retained more than 60% residual activity after incubation at 40°C for 1 h, while most GH11 *Ps*XLNs retained similar activity after 1 h at 30°C, with only *Ps*XlnJ and *Ps*XlnE being similar to GH10 XLNs (**Table 1**).

*Ps*XLNs displayed higher specific activity on BeWX than WAX, which is in line with commercially available XLNs [53] (**Table 1**). As an exception, *Ps*XlnI from GH11 was more active on WAX than on BeWX (**Table 1**). Additionally, among these different *Ps*XLNs, *Ps*XlnJ and *Ps*XlnG showed the lowest specific activity against xylan substrates.

A previous study described the transcriptome response of *P. subrubescens* to arabinoxylan-rich wheat bran and pectin-rich sugar beet pulp [42]. In this data set, the expression of the XLN-encoding genes from *P. subrubescens* grown on wheat bran was analyzed (**Suppl. Table 3**). The results revealed that all XLN-encoding genes were induced on wheat bran, with the highest induced expression being

Table 1 Biochemical properties of PsXLNs in this study.

GH family	Protein ID in JGI	Enzyme code	Optimum		Stability (>60%, 1h)		Specific activity (U/ μ mol)	
			pH	Temperature	pH	Temperature	BeWX	WAX
10	13043	<i>PsXlnA</i>	7	40	3.0-10.0	20-40	25301.5 \pm 4069.6	15093.7 \pm 562.8
	13445	<i>PsXlnB^a</i>	7	70	3.0-11.0	20-50	9576.5 \pm 613.5	5685.6 \pm 386.3
	12441	<i>PsXlnC</i>	5	70	5.0-10.0	20-40	3326.5 \pm 303.6	1997.1 \pm 36.2
	5870	<i>PsXlnD</i>	5	40	5.0-11.0	20-30	3242.7 \pm 302.6	1744.5 \pm 113.7
11	9034	<i>PsXlnE^a</i>	6	40	3.0-11.0	20-40	22280.5 \pm 2263.1	12650.6 \pm 1875.0
	1444	<i>PsXlnF</i>	6	70	6.0-8.0	20-30	2039.9 \pm 136.0	1069.5 \pm 210.8
	2115	<i>PsXlnG</i>	4	40	3.0-8.0	20-30	47.9 \pm 11.6	16.6 \pm 1.5
	3794	<i>PsXlnH</i>	6	40	3.0-11.0	20-30	5632.2 \pm 384.5	3006.4 \pm 195.9
	3721	<i>PsXlnI^a</i>	4	70	3.0-8.0	20-30	175.6 \pm 19.6	513.7 \pm 12.6
	7375	<i>PsXlnJ</i>	5	40	3.0-8.0	20-40	0.3 \pm 0.0	0.2 \pm 0.0

a, Enzyme possess a C-terminal CBM domain.

xlnA of GH10 and the lowest being *xlnJ* and *xlnG* in GH11. The low specific activity of *PsXlnJ* and *PsXlnG* against WAX may be related to the low expression of *xlnJ* and *xlnG* on wheat bran. However, it cannot be automatically concluded that the activity level of the enzymes and expression level of the corresponding genes correlate. *PsXlnA* and *PsXlnE*, as the most active enzymes towards WAX, showed similar activity levels, but the expression level of *xlnE* was about 4-fold higher than that of *xlnA*. Similar expression levels were observed in *xlnE* and *xlnF*, but the specific activity of *PsXlnE* towards WAX was 10-fold higher than that of *PsXlnF*. One possible reason for this could be that the media and culture conditions used do not reflect the heterogeneity and complexity of natural biotopes, and therefore may not reflect the natural response of *P. subrubescens*.

PsXLNs have diverse product profiles

To obtain an insight into the mode of action of XLNs, the product profiles of each independent XLN incubated with WAX for 16 h (end-point incubations) were analyzed by HPSEC-RI and HPAEC-PAD. *PsXlnJ* and *PsXlnG* were excluded from the study of product profile due to their low activity against WAX. HPSEC-RI analysis showed the disappearance of soluble polymers (WAX around 300 kDa) and the shift of molecular weight distribution below 1 kDa after 16 h incubation (**Suppl. Fig. 2**), indicating that all tested *PsXLNs* could extensively degrade WAX. To complement HPSEC-RI, HPAEC-PAD analysis demonstrated the presence of both linear XOS and substituted AXOS as a consequence of WAX degradation by XLNs (**Figure 2**).

X_2 , XA^3A^3X , $XA^3A^{2+3}XX$, and some unknown AXOS were clearly detected in all XLN-WAX digests, while the release of other annotated oligosaccharides (e.g., X_3 , $XA^{2+3}XX$) was enzyme dependent (**Figure 2**). A clear amount of D-xylose was also observed in XLN-WAX digests, except for the *PsXlnI*-WAX digest (**Figure 2B**). The comparison of product profiles is discussed in detail below.

In addition to the common products observed in all XLN-WAX digests, all GH10 XLNs could release $A^{2+3}XX$, A^3A^3X , and a trace amount of A^2XX from WAX. $XA^{2+3}XX$ was also detected in the *PsXlnA*- and *TmGH10*-WAX digests, which showed a more similar product profile to each other than to the others (**Figure 2A**). The annotated ‘peak 1’ refers to a double peak due to the co-elution of XA^3X , XA^2XX , A^3X . The double peak of ‘peak 1’ was clearly observed in *PsXlnA*- and *TmGH10*-WAX digests, whereas *PsXlnB*- and *PsXlnC*-WAX digests showed only a single peak. This suggests

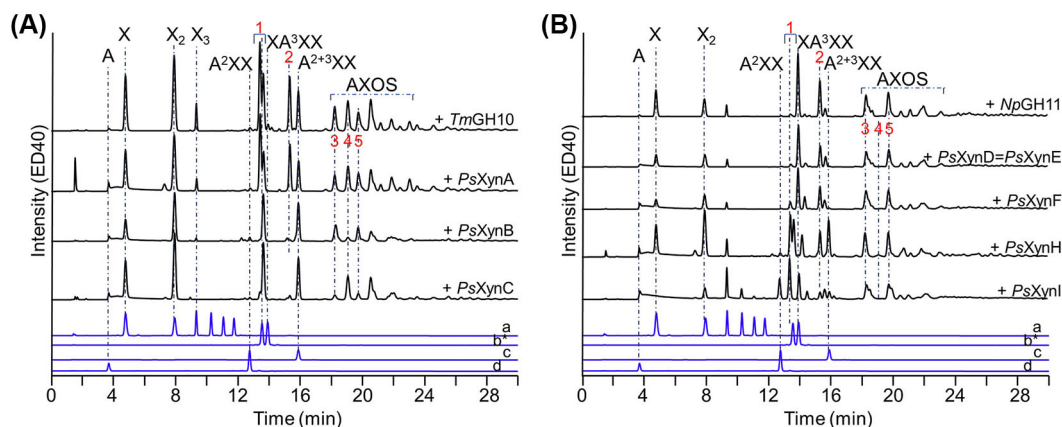


Figure 2 HPAEC elution patterns of wheat arabinoxyylan (WAX) digested with XLNs from GH10 (A) and GH11 (B) for 16 h. Arabinose (A), xylose (X), X_2 , X_3 , A^2XX , XA^3A^3X , $XA^{2+3}XX$ and $A^{2+3}XX$ were annotated based on analytical standards (bottom, A-B); other AXOS were annotated according to [36]; 1 = double peak due to co-elution of XA^3X , XA^2XX , A^3X (elution order); 2 = $XA^{2+3}XX$; 3 = XA^3A^3X ; 4 = A^3A^3X ; 5 = $XA^3A^{2+3}XX$; * XA^2XX eluted before XA^3A^3X .

that *PsXlnA*- and *TmGH10*-WAX digests contain more diverse components in ‘peak 1’. Notably, X_3 was present in *PsXlnA*-, *PsXlnB*- and *TmGH10*-WAX, expected to be an intermediate product, which is further hydrolyzed to X_2 and D-xylose upon depletion of longer substrates.

Unlike GH10-WAX digests, X_3 and $XA^{2+3}XX$ were observed in all tested GH11-WAX digests. XA^3XX was also found in GH11-WAX except in the *PsXlnH*-WAX digest (**Figure 2B**). Notably, some products were only present in specific XLN-WAX digests, such as A^2XX and $A^{2+3}XX$ in *PsXlnH*- and *PsXlnI*-WAX digests, and longer XOS (degree of polymerization, DP >3) in the *PsXlnI*-WAX digest. In addition, *PsXlnH*- and *PsXlnI*-WAX digests contain a larger amount of product in ‘peak 1’. Surprisingly, albeit only in trace amounts, A^3A^3X was present in the *PsXlnH*-WAX digest. In general, the product profiles of the *PsXlnE*-, *PsXlnD*- and *PsXlnF*-digests were similar to *NpGH11* [35], while those of the *PsXlnI*- and *PsXlnH*-digests were more diverse.

The presence of XOS with mono- L-arabinosyl substitution at the *O*-2 position (e.g., A^2XX , XA^2XX) in XLN-WAX digests was unexpected, as this substitution in WAX has been reported only associated together with *O*-3 substitutions to form doubly substituted D-xylosyl residual [35, 54]. The reason for the appearance of such products remains unclear. Further validation of the composition of the WAX used is needed.

The addition of PsABFs to PsXLNs-WAX reaction enhanced reducing sugar release

A previous study confirmed that multiple GH51, GH54 and GH62 ABFs from *P. subrubescens* have the ability to degrade WAX and release L-arabinose, to varying degrees [21]. To effectively degrade WAX present in dietary fiber, the cooperative effect of *PsXLNs* with different *PsABFs* was evaluated here.

The addition of *PsABFs* improved the hydrolysis of WAX by *PsXLNs* to some extent, as indicated by the higher reducing sugar levels released from WAX by the combination of enzymes than by a single enzyme, especially in pairs of GH51 or GH54 *PsABFs* with *PsXLNs* (**Figure 3, Suppl. Table 4**). Previous studies found that the simultaneous addition of XLNs with GH51 or GH54 ABFs enhanced polysaccharide hydrolysis, achieving the highest sugar release compared to any enzyme alone or sequential addition of the enzymes [55-57]. Relative to GH51 or GH54 *PsABFs*, the overall cooperation of GH62 *PsABFs* with *PsXLNs* was weaker and some of the combinations, e.g., $XlnA+A_xhC$, $XlnC+A_xhC$, $XlnI+A_xhA$, $XlnA+A_xhD$, $XlnI+A_xhD$, showed no or even a negative contribution in the reducing sugar release (**Figure 3, Suppl. Table 4**). This could be explained by the high preference of GH62 *PsABFs* towards WAX rather than AXOS [58]. The cooperation effect for GH62 *PsABFs* and *PsXLNs* may be improved by sequential addition of the enzymes.

Differences in the catalytic center of PsXLNs affect their hydrolysis behavior

The differences observed in the product profiles of different *PsXLNs* in hydrolysis of WAX may be due to differences in their catalytic domain. Superposition models of *PsXLNs* on available crystal structures complexed with XOS/AXOS indicates interesting features in the catalytic domain of models, which help to understand how the substrate binding cleft of *PsXLNs* is able to interact with WAX.

The superposition of structures of GH10 models and *SoXyn10A* (PDB: 1V6V) revealed that all GH10 *PsXLNs* can accommodate an L-arabinosyl side chain linked to *O*-3/*O*-2 at subsite -3 and +1, and *O*-3 at subsite -2, which support certain products (e.g., $A^{2+3}XX$, A^3A^3X , A^2XX) released from the non-reducing end of WAX by GH10 *PsXLNs* [9, 10] (**Figure 4A-D**). Additionally, the loop in the “N209-N217” region (based on *SoXyn10A*) is far from the catalytic center of the models, which allows for more space in the catalytic domain, indicating that longer XOS (e.g., X_3 , X_4) can be tolerated at the (+) subsite in catalytic cleft of GH10 *PsXLNs* [9] (**Figure 4B-D**). A previous study has reported that some XLNs lacking a +3 subsite can tolerate L-arabinosyl substituents at subsite +2, whereas XLNs with a +3 subsite cannot [10]. This inference may apply to GH10 models

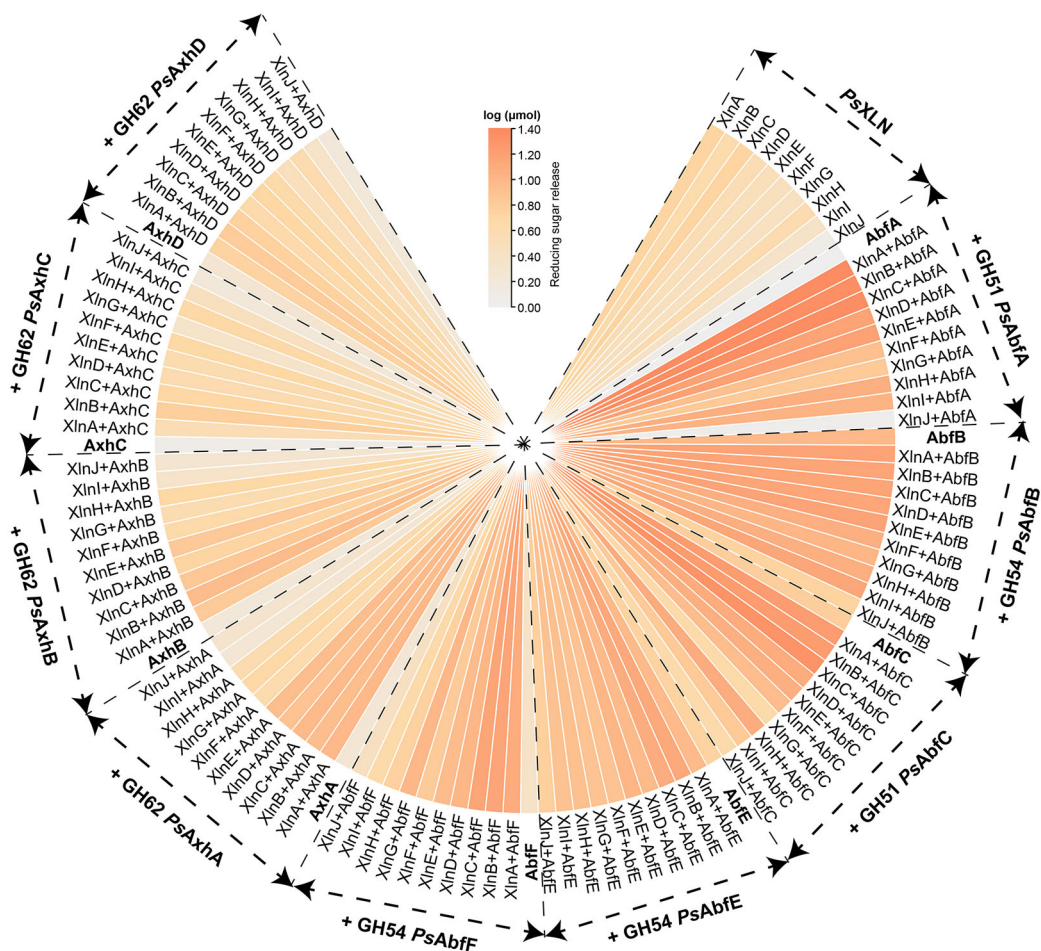


Figure 3 Heatmap of reducing sugar released from 24 h incubation of wheat arabinoxylan by *PsXLN* with/without the addition of *PsABF* from GH51, GH54 and GH62. *PsAbfA* (JGI: 7770) and *PsAbfC* (JGI: 1664) belong to GH51; *PsAbfB* (JGI: 1940), *PsAbfE* (JGI: 3364), and *PsAbfF* (JGI: 6724) belong to GH54; *PsAxA* (JGI: 12472), *PsAxB* (JGI: 12883), *PsAxC* (JGI: 3399), and *PsAxD* (JGI: 6027) belong to GH62.

other than *PsXlnA* which displays a wider cleft at subsite +2 due to a shorter loop in “N209-N217” region (**Figure 4B, 4E, Suppl. Fig. 3**). This structural feature might allow *PsXlnA* to accommodate substituted D-xylose at subsite +2 (**Figure 4B**), which could be an explanation for the presence of $XA^{2+3}XX$ in the *PsXlnA*-WAX digest (**Figure 2A**). Moreover, the high tolerance of the active site of *PsXlnA* to L-arabinosyl decorations leads to the production of more diverse AXOS than the other GH10 *PsXLNs* upon WAX hydrolysis (**Figures 2A, 4B**).

The comparison of structures between GH11 models and *SoXyn11A* (PDB: 7DFN) highlighted significant differences in the architecture of the catalytic center (termed “finger-thumb” regions) among the models, which is related to the different amino acids in the “G14-G23” and “R121-T132” regions and their arrangements in the catalytic domain [11] (**Figure 5, Suppl. Fig. 4**). These differences were expected to be a major factor to explain their diverse product profiles (**Figure 2B**). Generally, a partially closed conformation of “finger-thumb” regions is characteristic of the crystal structures of GH11 XLNs [59]. However, the corresponding region in several GH11 XLNs, i.e., *PsXlnI* and *PsXlnJ*, displays a fully open conformation, which makes the catalytic pocket more

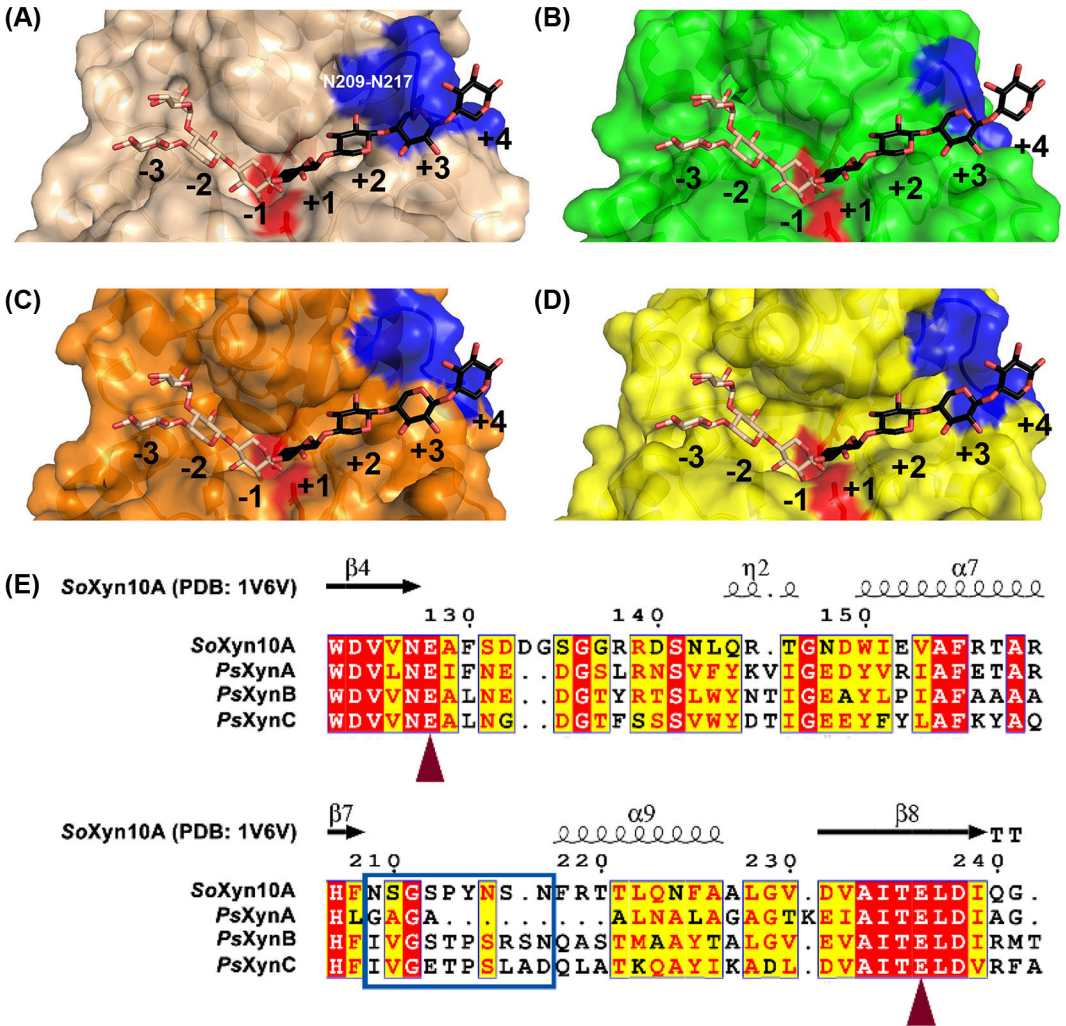


Figure 4 Substrate-binding cleft of GH10 PsXLNs. (A) *SoXyn10A* (PDB: 1V6V); (B) *PsXlnA*, (C) *PsXlnB*; (D) *PsXlnC*. Crystal structure and homology models superpositioned with XA³X (subsite -1 to -3, shown as white sticks) and X₄ (subsite +1 to +4, shown as black sticks); (E) alignment of GH10 PsXLNs (*PsXlnA-C*) with *Streptomyces olivaceoviridis* *SoXyn10A*. Catalytic residues and the region of main difference (‘N209-N217’ region based on *SoXyn10A*) in the catalytic domain were highlighted using red triangles and blue box in (E), respectively, and in red and blue in (A)-(D).

accessible (Figure 5B). This may contribute to direct interactions with the substrate and the unique degradation pattern of WAX by these enzymes.

Conclusions

Overall, the expansion of XLNs in the *P. subrubescens* genome was accompanied by functional diversity. Moreover, each independent XLN exhibited cooperative effects with *P. subrubescens* ABFs. These findings may reflect an evolutionary adaptation of this species that provides a wider enzymatic toolbox for synergistic degradation of hemicellulose in its natural habitat. Notably, the

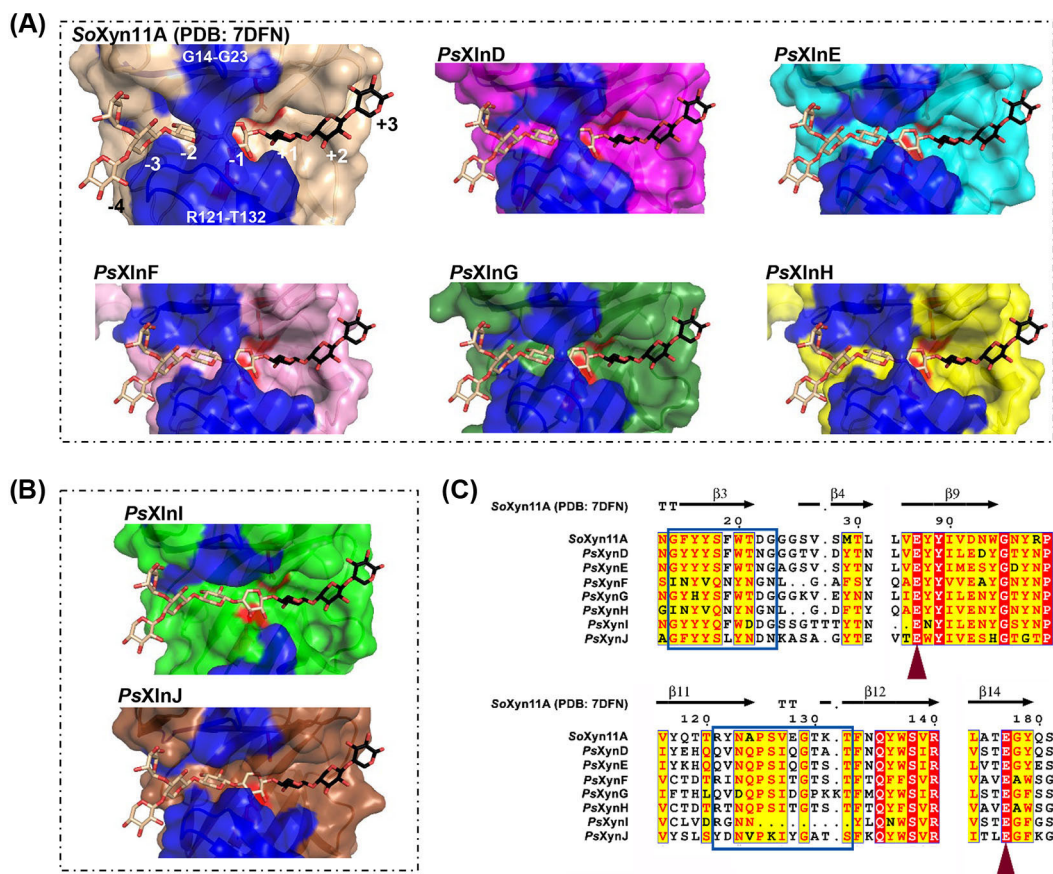


Figure 5 Substrate-binding cleft of GH11 PsXLNs. (A) Crystal structure and homology models with a partially closed pocket; (B) homology models with an open pocket. Crystal structure and homology models superpositioned with XA³XX (subsite -1 to -4, shown white sticks) and X₃ (subsite +1 to +3, shown black sticks); (C) alignment of GH11 PsXLNs (PsXlnD-J) with *Streptomyces olivaceoviridis* SoXyn11A. Catalytic residues and “finger-thumb” (“G14-G23” and “R121-T132” regions based on SoXyn11A) in the catalytic domain were pointed out using red triangle and blue box in (C), respectively, and highlighted in red and blue in (A)-(B).

diverse product patterns of WAX hydrolysis by *P. subrubescens* XLNs results from the different features in their catalytic domain. This finding contributes to the understanding of the interaction between enzyme and decorated xylan, and provides leads for enzyme engineering towards tailor-made XLNs.

Author contributions

RPdV conceived and supervised the overall project. RPdV and MAK designed the experiment. XL and DK conducted the experiments. XL wrote the original draft. All authors analyzed the data and commented on the manuscript.

Acknowledgements

The authors thank MSc Astrid Mueller for assistance with HPLC, Dr. Sumitha Krishnaswamyreddy for advice on structure analysis, and Dr. Adiphol Dilokpimol for suggestions on gene cloning. XL was supported by the China Scholarship Council for financial support (grant no. 201803250066).

References

1. Hansen, C. and A. Björkman, *The ultrastructure of wood from a solubility parameter point of view*. 1998.
2. Terrett, O.M., et al., *Molecular architecture of softwood revealed by solid-state NMR*. **Nature Communications**, 2019. 10(1): p. 1-11.
3. Amorim, C., et al., *From lignocellulosic residues to market: production and commercial potential of xylooligosaccharides*. **Biotechnology Advances**, 2019. 37(7): p. 107397.
4. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013*. **Nucleic Acids Research**, 2014. 42(D1): p. D490-D495.
5. Bastawde, K., *Xylan structure, microbial xylanases, and their mode of action*. **World Journal of Microbiology and Biotechnology**, 1992. 8(4): p. 353-368.
6. Paës, G., J.-G. Berrin, and J. Beaugrand, *GH11 xylanases: structure/function/properties relationships and applications*. **Biotechnology Advances**, 2012. 30(3): p. 564-592.
7. Pollet, A., J.A. Delcour, and C.M. Courtin, *Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families*. **Critical Reviews in Biotechnology**, 2010. 30(3): p. 176-191.
8. Capetti, C.C.d.M., et al., *Recent advances in the enzymatic production and applications of xylooligosaccharides*. **World Journal of Microbiology and Biotechnology**, 2021. 37(10): p. 1-12.
9. Fujimoto, Z., et al., *Crystal structures of decorated xylooligosaccharides bound to a family 10 xylanase from Streptomyces olivaceoviridis E-86*. **Journal of Biological Chemistry**, 2004. 279(10): p. 9606-9614.
10. Pell, G., et al., *The mechanisms by which family 10 glycoside hydrolases bind decorated substrates*. **Journal of Biological Chemistry**, 2004. 279(10): p. 9597-9605.
11. Fujimoto, Z., et al., *Structure-based substrate specificity analysis of GH11 xylanase from Streptomyces olivaceoviridis E-86*. **Applied Microbiology and Biotechnology**, 2021. 105(5): p. 1943-1952.
12. Hu, J. and J.N. Saddler, *Why does GH10 xylanase have better performance than GH11 xylanase for the deconstruction of pretreated biomass?* **Biomass and Bioenergy**, 2018. 110: p. 13-16.
13. Polizeli, M.d.L.T.d.M., et al., *Xylanases from fungi: properties and industrial applications*. **Applied Microbiology and Biotechnology**, 2005. 67(5): p. 577-591.
14. Li, X., et al., *Fungal xylanolytic enzymes: Diversity and applications*. **Bioresource Technology**, 2022. 344: p. 126290.
15. Dar, F.M. and P.M. Dar, *Fungal Xylanases for Different Industrial Applications*, in *Industrially Important Fungi for Sustainable Development*. **Springer**, 2021: p. 515-539.
16. Bhardwaj, N., K. Agrawal, and P. Verma, *Xylanases: An Overview of its Diverse Function in the Field of Biorefinery*. **Bioenergy Research: Commercial Opportunities & Challenges**, 2021: p. 295-317.
17. Aguilar-Pontes, M., et al., *The gold-standard genome of Aspergillus niger NRRL 3 enables a detailed view of the diversity of sugar catabolism in fungi*. **Studies in Mycology**, 2019. 92(1): p. 47-133.
18. Martinez, D., et al., *Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina)*. **Nature Biotechnology**, 2008. 26(5): p. 553-560.
19. Peng, M., et al., *The draft genome sequence of the ascomycete fungus Penicillium subrubescens reveals a highly enriched content of plant biomass related CAZymes compared to related fungi*. **Journal of Biotechnology**, 2017. 246: p. 1-3.
20. Linares, N.C., et al., *Recombinant production and characterization of six novel GH27 and GH36 α -galactosidases from Penicillium subrubescens and their synergism with a commercial mannanase during the hydrolysis of lignocellulosic biomass*. **Bioresource Technology**, 2020. 295: p. 122258.
21. Coconi Linares, N., et al., *Comparative characterization of nine novel GH51, GH54 and GH62 α -l-arabinofuranosidases from Penicillium subrubescens*. **FEBS Letters**, 2022. 596(3): p. 360-368.
22. Grigoriev, I.V., et al., *MycCosm portal: gearing up for 1000 fungal genomes*. **Nucleic Acids Research**, 2014. 42(D1): p. D699-D704.
23. Almagro Armenteros, J.J., et al., *SignalP 5.0 improves signal peptide predictions using deep neural networks*. **Nature Biotechnology**, 2019. 37(4): p. 420-423.
24. Katoh, K., J. Rozewicki, and K.D. Yamada, *MAFFT online service: multiple sequence alignment, interac-*

- tive sequence choice and visualization. **Briefings in Bioinformatics**, 2019. 20(4): p. 1160-1166.
25. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets*. **Molecular Biology and Evolution**, 2016. 33(7): p. 1870-1874.
 26. Van Campenhout, S. and G. Volckaert, *Differential expression of endo- β -1, 4-xylanase isoenzymes XI and X-II at various stages throughout barley development*. **Plant Science**, 2005. 169(3): p. 512-522.
 27. Chen, N.J. and R.E. Paull, *Endoxylanase expressed during papaya fruit ripening: purification, cloning and characterization*. **Functional Plant Biology**, 2003. 30(4): p. 433-441.
 28. Bih, F.Y., et al., *The predominant protein on the surface of maize pollen is an endoxylanase synthesized by a tapetum mRNA with a long 5' leader*. **Journal of Biological Chemistry**, 1999. 274(32): p. 22884-22894.
 29. Kunst, F., et al., *The complete genome sequence of the gram-positive bacterium Bacillus subtilis*. **Nature**, 1997. 390(6657): p. 249-256.
 30. Millward-Sadler, S., et al., *Evidence for a general role for high-affinity non-catalytic cellulose binding domains in microbial plant cell wall hydrolyses*. **Molecular Microbiology**, 1994. 11(2): p. 375-382.
 31. DeBoy, R.T., et al., *Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus*. **Journal of Bacteriology**, 2008. 190(15): p. 5455-5463.
 32. Li, X., et al., *Functional validation of two fungal subfamilies in carbohydrate esterase family 1 by biochemical characterization of esterases from uncharacterized branches*. **Frontiers in Bioengineering and Biotechnology**, 2020: p. 694.
 33. Lindsay, H., *A colorimetric estimation of reducing sugars in potatoes with 3, 5-dinitrosalicylic acid*. **Potato Research**, 1973. 16(3): p. 176-179.
 34. Van Gool, M., et al., *Performance of hemicellulolytic enzymes in culture supernatants from a wide range of fungi on insoluble wheat straw and corn fiber fractions*. **Bioresource Technology**, 2012. 114: p. 523-528.
 35. McCleary, B.V., et al., *Hydrolysis of wheat flour arabinoxylan, acid-debranched wheat flour arabinoxylan and arabino-xylo-oligosaccharides by β -xylanase, α -L-arabinofuranosidase and β -xylosidase*. **Carbohydrate Research**, 2015. 407: p. 79-96.
 36. Van Laere, K., et al., *Purification and mode of action of two different arabinoxylan arabinofuranohydrolases from Bifidobacterium adolescentis DSM 20083*. **Applied Microbiology and Biotechnology**, 1999. 51(5): p. 606-613.
 37. Gouet, P., et al., *ESPrpt: analysis of multiple sequence alignments in PostScript*. **Bioinformatics (Oxford, England)**, 1999. 15(4): p. 305-308.
 38. Li, X., et al., *Glycoside Hydrolase family 30 harbors fungal subfamilies with distinct polysaccharide specificities*. **New Biotechnology**, 2022. 67: p. 32-41.
 39. DeLano, W.L., *Pymol: An open-source molecular graphics tool*. CCP4 Newsl. **Protein Crystallography**, 2002. 40(1): p. 82-92.
 40. Chu, Y., et al., *Insights into the roles of non-catalytic residues in the active site of a GH10 xylanase with activity on cellulose*. **Journal of Biological Chemistry**, 2017. 292(47): p. 19315-19327.
 41. Wan, Q., et al., *X-ray crystallographic studies of family 11 xylanase Michaelis and product complexes: implications for the catalytic mechanism*. **Acta Crystallographica Section D: Biological Crystallography**, 2014. 70(1): p. 11-23.
 42. Dilokpimol, A., et al., *Penicillium subrubescens adapts its enzyme production to the composition of plant biomass*. **Bioresource Technology**, 2020. 311: p. 123477.
 43. Echeverría, V. and J. Eyzaguirre, *Penicillium purpurogenum produces a set of endoxylanases: identification, heterologous expression, and characterization of a fourth xylanase, XynD, a novel enzyme belonging to glycoside hydrolase family 10*. **Applied Biochemistry and Biotechnology**, 2019. 187(1): p. 298-309.
 44. Watanabe, M., et al., *Xylanase (GH11) from Acremonium cellulolyticus: homologous expression and characterization*. **AMB Express**, 2014. 4(1): p. 1-8.
 45. Laothanachareon, T., et al., *Synergistic action of recombinant accessory hemicellulolytic and pectinolytic enzymes to Trichoderma reesei cellulase on rice straw degradation*. **Bioresource Technology**, 2015. 198: p. 682-690.
 46. Sydenham, R., et al., *Cloning and enzymatic characterization of four thermostable fungal endo-1, 4- β -xylanases*. **Applied Microbiology and Biotechnology**, 2014. 98(8): p. 3613-3628.
 47. Takahashi, Y., H. Kawabata, and S. Murakami, *Analysis of functional xylanases in xylan degradation by Aspergillus niger E-1 and characterization of the GH family 10 xylanase XynVII*. **SpringerPlus**, 2013. 2(1): p. 1-11.
 48. Wang, J., et al., *Molecular cloning and heterologous expression of an acid-stable endoxylanase gene from*

- Penicillium oxalicum* in *Trichoderma reesei*. **Journal of Microbiology and Biotechnology**, 2013. 23(2): p. 251-259.
49. Xiong, K., et al., *Mutagenesis of N-terminal residues confer thermostability on a Penicillium janthinellum MA21601 xylanase*. **BMC Biotechnology**, 2019. 19(1): p. 1-9.
50. Yang, Y., et al., *Cooperation of hydrolysis modes among xylanases reveals the mechanism of hemicellulose hydrolysis by Penicillium chrysogenum P33*. **Microbial Cell Factories**, 2019. 18(1): p. 1-13.
51. Anthony, T., et al., *High molecular weight cellulase-free xylanase from alkali-tolerant Aspergillus fumigatus ARI*. **Enzyme and Microbial Technology**, 2003. 32(6): p. 647-654.
52. Taneja, K., S. Gupta, and R.C. Kuhad, *Properties and application of a partially purified alkaline xylanase from an alkalophilic fungus Aspergillus nidulans KK-99*. **Bioresource Technology**, 2002. 85(1): p. 39-42.
53. Mangan, D., et al., *Novel substrates for the automated and manual assay of endo-1, 4- β -xylanase*. **Carbohydrate Research**, 2017. 445: p. 14-22.
54. Sørensen, H.R., et al., *Enzymatic hydrolysis of wheat arabinoxylan by a recombinant "minimal" enzyme cocktail containing β -xylosidase and novel endo-1, 4- β -xylanase and α -L-arabinofuranosidase activities*. **Biotechnology Progress**, 2007. 23(1): p. 100-107.
55. De Wet, B.J., et al., *Characterization of a family 54 α -L-arabinofuranosidase from Aureobasidium pullulans*. **Applied Microbiology and Biotechnology**, 2008. 77(5): p. 975-983.
56. Tu, T., et al., *A GH51 α -L-arabinofuranosidase from Talaromyces leycettanus strain JCM12802 that selectively drives synergistic lignocellulose hydrolysis*. **Microbial Cell Factories**, 2019. 18(1): p. 1-12.
57. Yang, W., et al., *A novel bifunctional GH51 exo- α -l-arabinofuranosidase/endo-xylanase from Alicyclobacillus sp. A4 with significant biomass-degrading capacity*. **Biotechnology for Biofuels**, 2015. 8(1): p. 1-11.
58. Wilkens, C., et al., *GH62 arabinofuranosidases: structure, function and applications*. **Biotechnology Advances**, 2017. 35(6): p. 792-804.
59. Törrönen, A., A. Harkki, and J. Rouvinen, *Three-dimensional structure of endo-1, 4-beta-xylanase II from Trichoderma reesei: two conformational states in the active site*. **The EMBO Journal**, 1994. 13(11): p. 2493-2501.

Supplementary material

Suppl. Table 1 Amino acid sequences used for the GH10 and GH11 phylogeny. The outgroup sequences are highlighted in yellow. **Available upon request from author.**

Suppl. Table 2 Primers used for cloning in this study.

Suppl. Table 3 Expression of *PsXLN*-encoding genes on wheat bran (data based on previous study [42]).

Suppl. Table 4 Reducing sugar released from 24 h incubation of wheat arabinoxylan by *PsXLN* with/without the addition of *PsABFs* from GH51, GH54 and GH62. *PsAbfA* (JGI: 7770) and *PsAbfC* (JGI: 1664) belong to GH51; *PsAbfB* (JGI: 1940), *PsAbfE* (JGI: 3364), and *PsAbfF* (JGI: 6724) belong to GH54; *PsAxB* (JGI: 12472), *PsAxB* (JGI: 12883), *PsAxB* (JGI: 3399), and *PsAxB* (JGI: 6027) belong to GH62.

Suppl. Table 5 Summary of the validation of closest structural models of *PsXLNs*.

Suppl. Fig. 1 Analysis of purified recombinant *PsXLNs* with 12% (w/v) tris-glycine SDS-PAGE.

Suppl. Fig. 2 Molecular weight distributions of wheat arabinoxylan (WAX) digested by 16 h incubation with XLNs from GH10 (A) and GH11 (B) and analyzed by HPSEC-RI.

Suppl. Fig. 3 Alignment of GH10 *PsXLNs* (*PsXlnA-C*) with *Streptomyces olivaceoviridis* SoXyn10A. Catalytic residues and the region of main difference ("N209-N217" region based on SoXyn10A) in catalytic domain were pointed out using red triangle and blue box, respectively.

Suppl. Fig. 4 Alignment of GH11 *Ps*XLNs (*Ps*XlnD-J) with *Streptomyces olivaceoviridis* *So*Xyn11A. Catalytic residues and “finger-thumb” (“G14-G23” and “R121-T132” regions based on *So*Xyn11A) in catalytic domain were pointed out using red triangle and blue box, respectively.

Suppl. Table 2 Primers used for cloning in this study.

Family	Protein id	Restriction enzymes	Forward primer	Reverse primer
GH10	13445	EcoRI/NotI	AATCAGAAT GAATTC GCTGGCTTGAACACTGTGTC	ATTCTGAAT GGGCCG CCAGACACTGGGAGTACCAGGC
	12441	EcoRI/NotI	AATCAGAAT GAATTC GCACCTCCATCCACCTCAC	ATTCTGAAT GGGCCG CGGCACACACTGCAAGGCTTTCC
	13043	EcoRI/NotI	AATCAGAAT GAATTC AGCCCTGTCGAGATCGACTCC	ATTCTGAAT GGGCCG CAAGCGCACTGGGATAGCGTTG
GH11	2115	EcoRI/NotI	AATCAGAAT GAATTC GCTCCCACTGAAATCAATGTGGCAAG	ATTCTGAAT GGGCCG CCCAAAAGACATATGAGCGCGTCC
	7375	NotI/XbaI	AATCAGAAT GGGGCCG CAAGCCCTCAGCACCCATGAAAAG	ATTCTGAAT TCTAGAT TCCGCACAGTAAAATCAAGATAACCCCTTGC
GH12	1444	EcoRI/NotI	AATCAGAAT GAATTC GCCCCCTGCCACCCGATCTAGTG	ATTCTGAAT GGGCCG AGCTGAGACCCGTGACACTAGCGGTG
	3721	EcoRI/NotI	AATCAGAAT GAATTC AGCCCTGCGGCAGCAC	ATTCTGAAT GGGCCG CGCGGTTGAAGAGAAAAGGTGCCAG
	3794	EcoRI/NotI	AATCAGAAT GAATTC GCTCCTGGAACCGAACCTGG	ATTCTGAAT GGGCCG CTGAAAATATCGTAAACCCCTGGCGCTG
	5870	EcoRI/NotI	AATCAGAAT GAATTC GTTCCCGCGAATGGAACAAGC	ATTCTGAAT GGGCCG CGAGAAAACAGTATGGAAGAAGAGCCCGCTG
	9034	EcoRI/NotI	AATCAGAAT GAATTC CTGCCTGGCGACATGTCCG	ATTCTGAAT GGGCCG CGGAGGCACCTGCGAGTACCACG

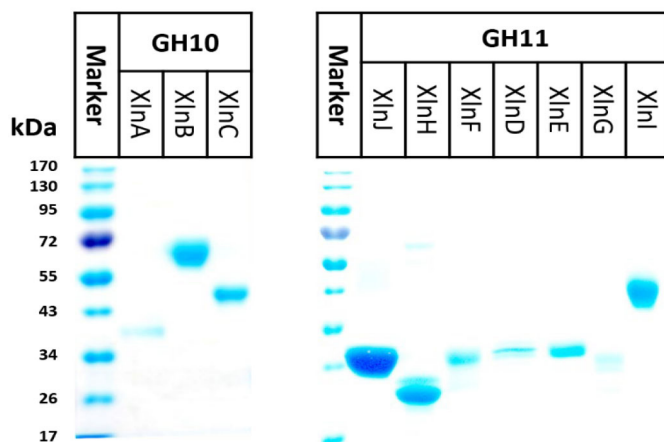
The recognition sequences of restriction endonuclease (EcoRI/NotI or NotI/XbaI) were highlighted in red.

Suppl. Table 3 Expression of *PsXLN*-encoding genes on wheat bran (data based on previous study [42]).

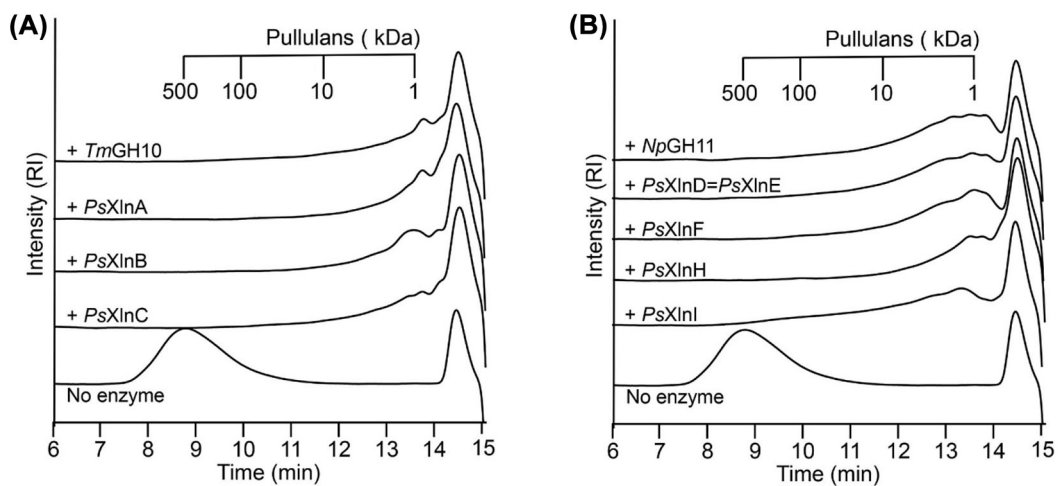
CAZy family	JGI protein ID	Enzyme name	Expression level (FPKM)	SD
GH10	13445	<i>PsXlnB</i>	2107.0	204.9
	12441	<i>PsXlnC</i>	172.5	80.0
	13043	<i>PsXlnA</i>	38888.3	2080.0
GH11	5870	<i>PsXlnD</i>	2634.3	276.8
	9034	<i>PsXlnE</i>	8963.1	58.4
	3794	<i>PsXlnH</i>	580.5	5.3
	3721	<i>PsXlnI</i>	608.5	26.7
	1444	<i>PsXlnF</i>	9957.0	878.5
	7375	<i>PsXlnJ</i>	7.4	2.6
	2115	<i>PsXlnG</i>	2.4	2.1

Suppl. Table 4 Reducing sugar released from 24 h incubation of wheat arabinoxylan by *Ps*XLN with/ without the addition of *Ps*ABFs from GH51, GH54 and GH62. *Ps*AbfA (JGI: 7770) and *Ps*AbfC (JGI: 1664) belong to GH51; *Ps*AbfB (JGI: 1940), *Ps*AbfE (JGI: 3364), and *Ps*AbfF (JGI: 6724) belong to GH54; *Ps*AxhA (JGI: 12472), *Ps*AxhB (JGI: 12883), *Ps*AxhC (JGI: 3399), and *Ps*AxhD (JGI: 6027) belong to GH62.

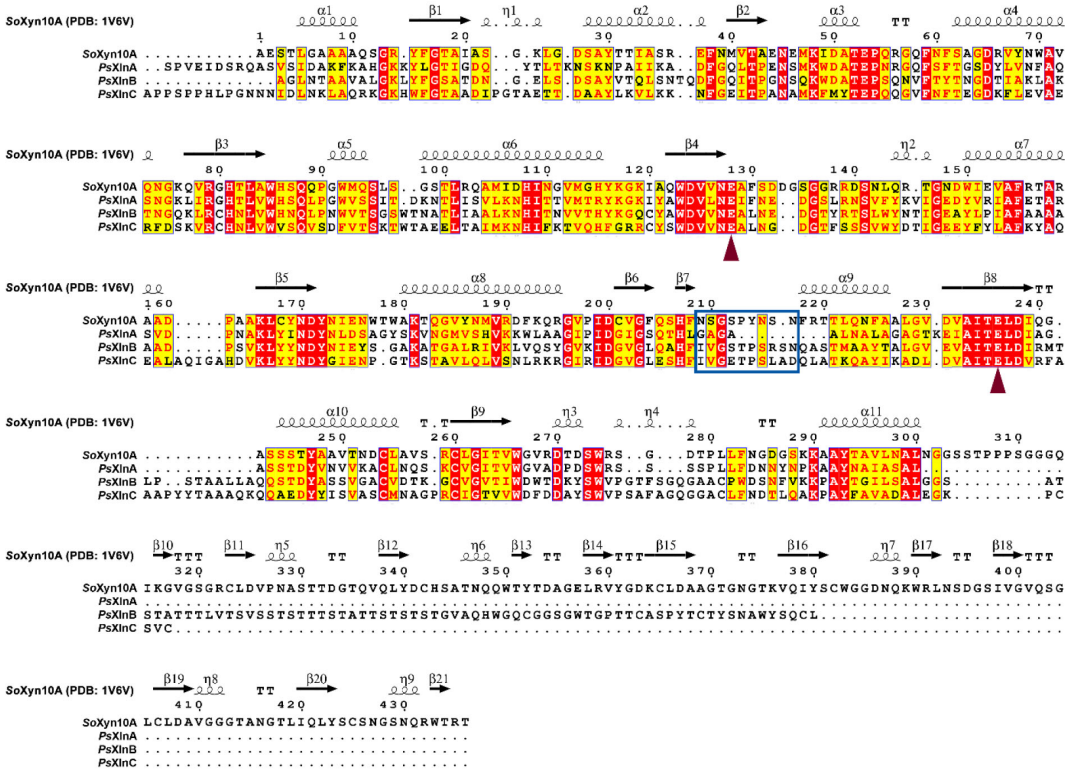
Enzyme name	Reducing sugar release (μmol)	Enzyme name	Reducing sugar release (μmol)	Enzyme name	Reducing sugar release (μmol)
<i>Ps</i> XlnA	0.666	<i>Ps</i> XlnE	0.421	<i>Ps</i> XlnH	0.527
<i>Ps</i> XlnB	0.548	<i>Ps</i> XlnF	0.387	<i>Ps</i> XlnI	0.304
<i>Ps</i> XlnC	0.664	<i>Ps</i> XlnG	0.357	<i>Ps</i> XlnJ	0.072
<i>Ps</i> XlnD	0.475	<i>Ps</i> AbfE (JGI: 3364)	0.679	<i>Ps</i> AxhB (JGI: 12883)	0.153
<i>Ps</i> AbfA (JGI: 7770)	0.017	<i>Ps</i> XlnA+ <i>Ps</i> AbfE	0.967	<i>Ps</i> XlnA+ <i>Ps</i> AxhB	0.768
<i>Ps</i> XlnA+ <i>Ps</i> AbfA	1.602	<i>Ps</i> XlnB+ <i>Ps</i> AbfE	1.178	<i>Ps</i> XlnB+ <i>Ps</i> AxhB	0.945
<i>Ps</i> XlnB+ <i>Ps</i> AbfA	1.559	<i>Ps</i> XlnC+ <i>Ps</i> AbfE	1.128	<i>Ps</i> XlnC+ <i>Ps</i> AxhB	0.807
<i>Ps</i> XlnC+ <i>Ps</i> AbfA	1.581	<i>Ps</i> XlnD+ <i>Ps</i> AbfE	0.958	<i>Ps</i> XlnD+ <i>Ps</i> AxhB	0.661
<i>Ps</i> XlnD+ <i>Ps</i> AbfA	1.251	<i>Ps</i> XlnE+ <i>Ps</i> AbfE	1.006	<i>Ps</i> XlnE+ <i>Ps</i> AxhB	0.788
<i>Ps</i> XlnE+ <i>Ps</i> AbfA	1.275	<i>Ps</i> XlnF+ <i>Ps</i> AbfE	0.860	<i>Ps</i> XlnF+ <i>Ps</i> AxhB	0.587
<i>Ps</i> XlnF+ <i>Ps</i> AbfA	0.898	<i>Ps</i> XlnG+ <i>Ps</i> AbfE	0.859	<i>Ps</i> XlnG+ <i>Ps</i> AxhB	0.505
<i>Ps</i> XlnG+ <i>Ps</i> AbfA	0.781	<i>Ps</i> XlnH+ <i>Ps</i> AbfE	0.883	<i>Ps</i> XlnH+ <i>Ps</i> AxhB	0.626
<i>Ps</i> XlnH+ <i>Ps</i> AbfA	1.087	<i>Ps</i> XlnI+ <i>Ps</i> AbfE	0.873	<i>Ps</i> XlnI+ <i>Ps</i> AxhB	0.292
<i>Ps</i> XlnI+ <i>Ps</i> AbfA	1.039	<i>Ps</i> XlnJ+ <i>Ps</i> AbfE	0.719	<i>Ps</i> XlnJ+ <i>Ps</i> AxhB	0.226
<i>Ps</i> XlnJ+ <i>Ps</i> AbfA	0.030	<i>Ps</i> AbfF (JGI: 6724)	0.307	<i>Ps</i> AxhC (JGI: 3399)	0.056
<i>Ps</i> AbfB (JGI: 1940)	0.976	<i>Ps</i> XlnA+ <i>Ps</i> AbfF	1.147	<i>Ps</i> XlnA+ <i>Ps</i> AxhC	0.468
<i>Ps</i> XlnA+ <i>Ps</i> AbfB	1.246	<i>Ps</i> XlnB+ <i>Ps</i> AbfF	1.232	<i>Ps</i> XlnB+ <i>Ps</i> AxhC	0.728
<i>Ps</i> XlnB+ <i>Ps</i> AbfB	1.234	<i>Ps</i> XlnC+ <i>Ps</i> AbfF	1.220	<i>Ps</i> XlnC+ <i>Ps</i> AxhC	0.637
<i>Ps</i> XlnC+ <i>Ps</i> AbfB	1.238	<i>Ps</i> XlnD+ <i>Ps</i> AbfF	1.006	<i>Ps</i> XlnD+ <i>Ps</i> AxhC	0.524
<i>Ps</i> XlnD+ <i>Ps</i> AbfB	1.218	<i>Ps</i> XlnE+ <i>Ps</i> AbfF	0.944	<i>Ps</i> XlnE+ <i>Ps</i> AxhC	0.522
<i>Ps</i> XlnE+ <i>Ps</i> AbfB	1.246	<i>Ps</i> XlnF+ <i>Ps</i> AbfF	0.766	<i>Ps</i> XlnF+ <i>Ps</i> AxhC	0.618
<i>Ps</i> XlnF+ <i>Ps</i> AbfB	1.084	<i>Ps</i> XlnG+ <i>Ps</i> AbfF	0.984	<i>Ps</i> XlnG+ <i>Ps</i> AxhC	0.279
<i>Ps</i> XlnG+ <i>Ps</i> AbfB	1.003	<i>Ps</i> XlnH+ <i>Ps</i> AbfF	0.695	<i>Ps</i> XlnH+ <i>Ps</i> AxhC	0.602
<i>Ps</i> XlnH+ <i>Ps</i> AbfB	1.099	<i>Ps</i> XlnI+ <i>Ps</i> AbfF	0.587	<i>Ps</i> XlnI+ <i>Ps</i> AxhC	0.386
<i>Ps</i> XlnI+ <i>Ps</i> AbfB	1.205	<i>Ps</i> XlnJ+ <i>Ps</i> AbfF	0.363	<i>Ps</i> XlnJ+ <i>Ps</i> AxhC	0.160
<i>Ps</i> XlnJ+ <i>Ps</i> AbfB	0.686	<i>Ps</i> AxhA (JGI: 12472)	0.198	<i>Ps</i> AxhD (JGI: 6027)	0.217
<i>Ps</i> AbfC (JGI: 1664)	0.731	<i>Ps</i> XlnA+ <i>Ps</i> AxhA	0.971	<i>Ps</i> XlnA+ <i>Ps</i> AxhD	0.679
<i>Ps</i> XlnA+ <i>Ps</i> AbfC	1.344	<i>Ps</i> XlnB+ <i>Ps</i> AxhA	0.915	<i>Ps</i> XlnB+ <i>Ps</i> AxhD	0.755
<i>Ps</i> XlnB+ <i>Ps</i> AbfC	1.352	<i>Ps</i> XlnC+ <i>Ps</i> AxhA	0.969	<i>Ps</i> XlnC+ <i>Ps</i> AxhD	0.696
<i>Ps</i> XlnC+ <i>Ps</i> AbfC	1.485	<i>Ps</i> XlnD+ <i>Ps</i> AxhA	0.861	<i>Ps</i> XlnD+ <i>Ps</i> AxhD	0.567
<i>Ps</i> XlnD+ <i>Ps</i> AbfC	1.146	<i>Ps</i> XlnE+ <i>Ps</i> AxhA	0.914	<i>Ps</i> XlnE+ <i>Ps</i> AxhD	0.644
<i>Ps</i> XlnE+ <i>Ps</i> AbfC	1.092	<i>Ps</i> XlnF+ <i>Ps</i> AxhA	0.542	<i>Ps</i> XlnF+ <i>Ps</i> AxhD	0.576
<i>Ps</i> XlnF+ <i>Ps</i> AbfC	0.932	<i>Ps</i> XlnG+ <i>Ps</i> AxhA	0.614	<i>Ps</i> XlnG+ <i>Ps</i> AxhD	0.410
<i>Ps</i> XlnG+ <i>Ps</i> AbfC	0.586	<i>Ps</i> XlnH+ <i>Ps</i> AxhA	0.518	<i>Ps</i> XlnH+ <i>Ps</i> AxhD	0.532
<i>Ps</i> XlnH+ <i>Ps</i> AbfC	1.122	<i>Ps</i> XlnI+ <i>Ps</i> AxhA	0.198	<i>Ps</i> XlnI+ <i>Ps</i> AxhD	0.301
<i>Ps</i> XlnI+ <i>Ps</i> AbfC	0.784	<i>Ps</i> XlnJ+ <i>Ps</i> AxhA	0.256	<i>Ps</i> XlnJ+ <i>Ps</i> AxhD	0.161
<i>Ps</i> XlnJ+ <i>Ps</i> AbfC	0.569				



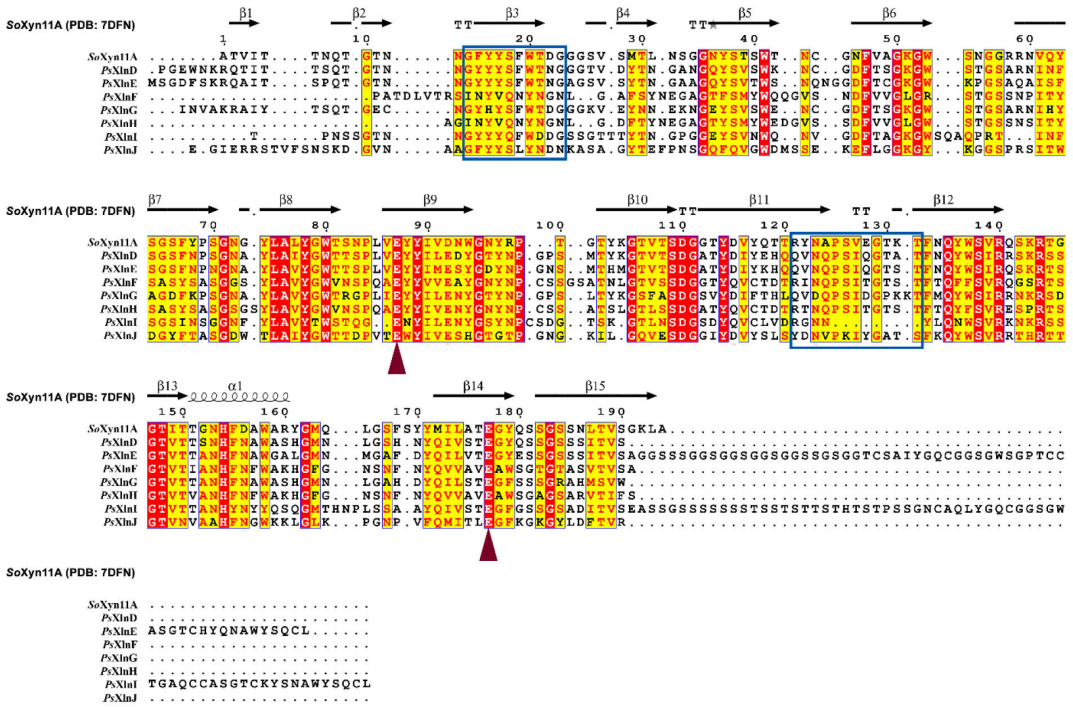
Suppl. Fig. 1 Analysis of purified recombinant *Ps*XLNs with 12% (w/v) tris–glycine SDS-PAGE.



Suppl. Fig. 2 Molecular weight distributions of wheat arabinoxylan (WAX) digested by 16 h incubation with XLNs from GH10 (A) and GH11 (B) and analyzed by HPSEC-RI.



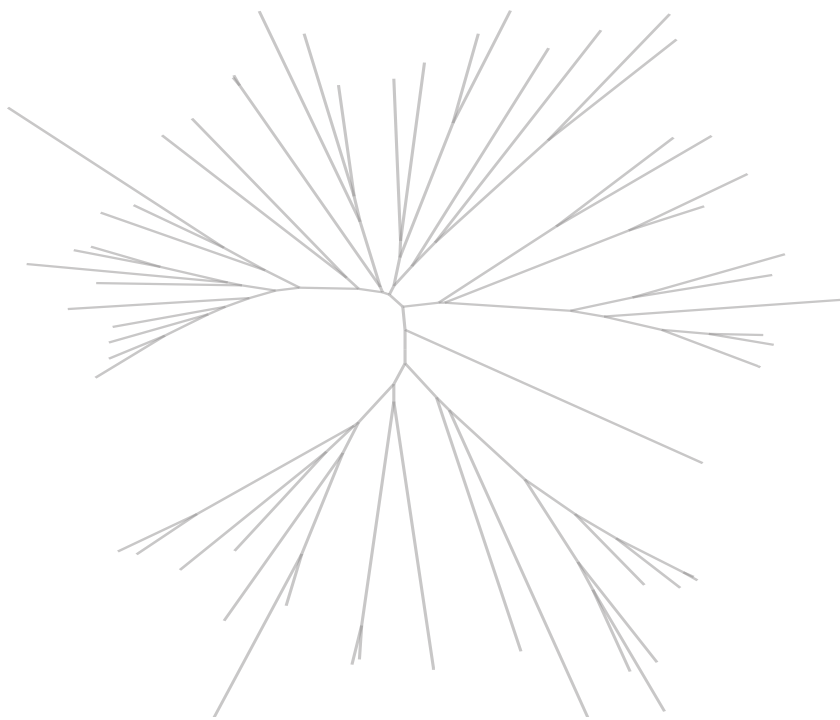
Suppl. Fig. 3 Alignment of GH10 *PsXLNs* (*PsXlnA-C*) with *Streptomyces olivaceoviridis* *SoXyn10A*. Catalytic residues and the region of main difference (“N209-N217” region based on *SoXyn10A*) in catalytic domain were pointed out using red triangle and blue box, respectively.



Suppl. Fig. 4 Alignment of GH11 *PsXLNs* (*PsXlnD*-*J*) with *Streptomyces olivaceoviridis* *SoXyn11A*. Catalytic residues and “finger-thumb” (“G14-G23” and “R121-T132” regions based on *SoXyn11A*) in catalytic domain were pointed out using red triangle and blue box, respectively.

章节六

真菌GH44家族的木聚糖酶独属于担子菌门，且呈现独特的木葡聚糖裂解模式



Chapter 6

Fungal Glycoside Hydrolase family 44 xyloglucanases are restricted to the phylum Basidiomycota and show a distinct xyloglucan cleavage pattern

This chapter was published in iScience

Sun, Peicheng, Xinxin Li, Adiphol Dilokpimol, Bernard Henrissat, Ronald P. de Vries, Mirjam A. Kabel, and Miia R. Mäkelä

Volume 25, 21 January 2022, 103666

<https://doi.org/10.1016/j.isci.2021.103666>

Abstract

Xyloglucan is a prominent matrix heteropolysaccharide binding to cellulose microfibrils in primary plant cell walls. Hence, the hydrolysis of xyloglucan facilitates the overall lignocellulosic biomass degradation. Xyloglucanases (XEGs) are key enzymes classified in several Glycoside Hydrolase (GH) families. So far, family GH44 has been shown to contain bacterial XEGs only. Detailed genome analysis revealed GH44 members in fungal species from the phylum Basidiomycota, but not in other fungi, which we hypothesized to also be XEGs. Two GH44 enzymes from *Dichomitus squalens* and *Pleurotus ostreatus* were heterologously produced and characterized. They exhibited XEG activity and displayed a hydrolytic cleavage pattern different from that observed in fungal XEGs from other GH families. Specifically, the fungal GH44 XEGs were not hindered by substitution of neighboring D-glucosyl units and generated various “XXXG-type”, “GXXX(G)-type” and “XXX-type” oligosaccharides. Overall, these fungal GH44 XEGs represent a novel class of enzymes for plant biomass conversion and valorization.

Introduction

Xyloglucan is present in all vascular plants, in particular within the primary cell walls, and is a major component of dicots (up to 25% of total dry weight) and non-commelinid monocots including conifers (10%) [1]. The rigidity of plant cell walls is largely driven by loosening and tightening of cellulose microfibrils, in which xyloglucan plays a key role owing to its binding ability to cellulose by surface adsorption and chain intercalation [2, 3]. In addition, xyloglucan is found as a major seed-storage polysaccharide in a number of terrestrial plants including tamarind and nasturtium [4]. Enzymatic hydrolysis of xyloglucan not only facilitates the access of cellulases to cellulose but also contributes to complete saccharification of agro-industrial side streams to drive the production of biofuels and biochemicals [5, 6]. Furthermore, soluble xyloglucan and its hydrolysis products can serve as relevant food and feed additives, for example, to reduce calorie content, to increase the texture of food products, or due to their prebiotic behavior [7-9].

Xyloglucan has a substituted β -(1 \rightarrow 4)-D-glucan backbone, of which the exact type and degree of substitution varies for different plant sources and tissues [1]. Still, most xyloglucan structures have been shown as the “XXGG-type” and “XXXG-type” [10], in which G stands for a D-glucosyl residue and X for a D-glucosyl residue substituted with an α -(1 \rightarrow 6)-D-xylosyl unit [11]. In addition, D-xylosyl residues can be further substituted by β -(1 \rightarrow 2)-D-galactosyl units (coded L) or, to a lesser extent, with α -(1 \rightarrow 2)-L-arabinosyl units (coded S). The D-galactosyl residues can be fucosylated at the O-2 position (coded F). Other less common substitutions are denoted with other designations described elsewhere in detail [4, 11].

To date, characterized xyloglucanases (XEGs) have been found in glycoside hydrolase (GH) families GH5, GH9, GH12, GH16, GH44 and GH74 in Carbohydrate-Active enZYme (CAZy) database (<http://www.cazy.org>) [12]. Fungal XEGs are, however, only found in families GH5 [13], GH12, and GH74 [14]. The so-far only characterized fungal GH5 XEG has been shown to be an endo-processive-type of XEG [13]. Fungal GH12 XEGs have been shown to catalyze the hydrolysis of the xyloglucan backbone in an endo-mode, cleaving the glycosidic linkages at the reducing end side of unsubstituted D-glucosyl units, releasing “XXXG-type” oligosaccharides [15-17]. Fungal GH5 and GH74 XEGs have also been shown to display an endo-type of hydrolysis of the xyloglucan backbone, cleaving at the “reducing end” side of unsubstituted D-glucosyl units like the GH12 XEGs [16, 17]. However, in contrast to the GH12 XEGs, the GH5 and GH74 XEGs display a processive mode of action [18-21]. As a result, from the early stage of hydrolysis, XXXG- and XLG-building blocks are released by GH74 XEGs, which can even be further degraded into XX, XG and LG [19-22]. GH5 XEG generated XXXG-type oligosaccharides, in addition to several non-XXXG-type oligosaccharides such as X, XG, XX, LG, XL/LX, XXG and XXX [13]. Detailed genome analysis revealed the presence of basidiomycete proteins in enzyme family GH44 of the CAZy database, for which only bacterial endoglucanases and XEGs have been reported so far [23]. Interestingly, no members from other fungal phyla were present in GH44. In this study, we evaluated whether the basidiomycete members of GH44 are also XEGs and compared their product profile with that of XEGs of families GH12 and GH74.

Materials and methods

Microbial strains and growth conditions

E. coli DH5 α competent cells (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA, USA) were employed to propagate pPicZaA-xegA^{Ds} and pPicZaA-xegA^{Po} plasmids and the transformed cells were grown in low-salt Luria Bertani medium supplemented with 25 μ g/mL Zeocin at 37°C. *P. pastoris* strain X-33 (Invitrogen) was transformed with pPicZaA-xegA^{Ds} and pPicZaA-xegA^{Po} plasmids and the positive transformants were selected on agar plates containing 1% yeast extract, 2% peptone, 2% glucose, 1 M sorbitol, 2% agar, and 100 μ g/mL Zeocin. For production of recombinant

proteins, the selected *P. pastoris* transformants were grown in Buffered Glycerol Complex medium and induced in Buffered Methanol Complex medium at 30°C [24].

Bioinformatics

The amino acid sequences of 190 GH44 candidates were retrieved from the JGI MycoCosm portal (<https://mycocosm.jgi.doe.gov/mycocosm/home>). After removing sequences from unpublished genomes and keeping only one strain per species, the sequences were aligned using Multiple Alignment using Fast Fourier Transform (MAFFT) [25]. After manually filtering and correcting the sequences for gene model errors, 57 sequences from 42 basidiomycete species were included in phylogenetic analysis (**Suppl. Table 1**). Because characterized bacterial GH44 enzymes appeared to be too distant to be used as an outgroup, unrooted Maximum Likelihood, Minimum Evolution (ME) and Neighbor Joining (NJ) trees were computed using 500 bootstraps in MEGA7 [26] to reveal the evolutionary relationship between the fungal GH44 sequences.

Cloning, protein production and purification of the selected candidates

The selected candidate genes (**Table 1**) without predicted signal peptides and introns were codon optimized and synthesized into pPicZαA plasmid (Genscript Biotech, Leiden, The Netherlands), which were then chemically transformed into *E. coli* DH5α and propagated. The plasmids pPicZαA-*xegA*^{Ds} and pPicZαA-*xegA*^{Po} were extracted and linearized with *PmeI* (New England Biolabs, Ipswich, MA) and subsequently transformed into *P. pastoris* strain X-33 by electroporation. The positive colonies were selected by colony Western Blot using 6x-His Tag Monoclonal Antibody conjugated to alkaline phosphatase (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA, USA) and detected with a chromogenic substrate 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (BCIP/NBT) [27].

The selected *P. pastoris* transformants were cultivated for the protein production according to [24]. After four days of induction, culture supernatants were harvested (8000 g, 4°C, 20 min) and purified using ÄKTA FPLC device (GE Healthcare Life Sciences, Uppsala, Sweden) as described earlier [24]. The purified proteins were stored at 4°C prior to further analysis.

Molecular mass and concentration of the purified enzymes were evaluated by SDS-PAGE and ImageJ program, respectively [27]. Deglycosylation was performed using endoglycosidase H (Endo H; New England Biolabs, Ipswich, MA) following the manufacturer's recommendation.

Enzyme substrate screening

Different substrates including TXG, Avicel[®]PH-101, galactomannan and beechwood xylan were suspended in 50 mM sodium acetate buffer (pH 5.0) to a final concentration of 2 mg/mL. *DsXegA* and *PoXegA* were added to a final protein concentration of 100 µg/mL. The digestion (reaction volume of 300 µL) was incubated at 40°C with shaking (Multitron shaking incubators, Infors HT, Bottmingen, Switzerland) at 110 rpm for 24 h. Control reactions were performed without the addition of enzymes. All digestions were performed in duplicate. The reactions were stopped by heating at 97°C for 10 min. Afterwards, the digests were centrifuged at 22000 g for 20 min, and the clear supernatants were collected and analyzed by high-performance anion exchange chromatography with HPAEC-PAD after diluting ten times.

Temperature and pH profiles

The temperature and pH optima of GH44 enzymes were determined at 20–80°C in 100 mM phosphate buffer, pH 5.0, or at 40°C in 100 mM Britton-Robinson buffer from pH 2.0 to pH 10.0 [28], respectively. The thermal and pH stability were determined by measuring the residual enzyme activity at 37°C after 1 h incubation at 20–70°C in 100 mM phosphate buffer (pH 5.0 for *DsXegA*

and pH 6.0 for *PoXegA*), or at 40°C in 100 mM Britton-Robinson buffer from pH 2.0 to pH 10.0. All digestions were performed in duplicate using TXG as a substrate.

Enzymatic digestion of xyloglucan substrates

Xyloglucan substrates (TXG or BCXG) were dissolved in 50 mM ammonium acetate buffer (pH 5.0) to a final concentration of 2 mg/mL. *DsXegA*, *PoXegA*, *AnXegA* and *AnXegB* were dosed in a final protein concentration of 100 µg/mL. The digestion (reaction volume of 300 µL) was incubated in an Eppendorf ThermoMixer® Comfort at 40°C with shaking at 800 rpm for 24 h. Control reactions were performed without the addition of enzymes. All digestions were performed in duplicate. The reactions were stopped by incubating at 97°C for 10 min. Afterwards, the digests were centrifuged at 22000 g for 20 min, and the clear supernatants were collected and stored at -20°C until further usage. Previous well characterized *NcLPMO9C*- and GH5-TXG digests were used as reference mixtures [29]. The TXG oligosaccharides released by bacterial GH5 are further referred as the “reference TXG oligosaccharides”. *DsXegA*-, *PoXegA*-, *AnXegA*- and *AnXegB*-TXG digests were analyzed by HPAEC-PAD after diluting 20 times, and by HILIC-ESI-CID-MS/MS². The TXG oligosaccharide standards were also analyzed by HPAEC-PAD (50 µg/mL) and HILIC-ESI-CID-MS/MS² (100 µg/mL). The *NcLPMO9C*-TXG digest and reference TXG oligosaccharides were analyzed by HPAEC-PAD after diluting five and ten times, respectively.

HPAEC-PAD analysis for substrate screening and for xyloglucan oligosaccharide profiling

For the substrate screening and the xyloglucan oligosaccharide profiling, TXG and other carbohydrates corresponding digests were analyzed by HPAEC-PAD on an ICS5000 (Thermo Scientific, Waltham, MA, USA) system. Instrument settings, column, mobile phases and elution program have been described previously [29].

HILIC-ESI-CID-MS/MS² for structural elucidation of xyloglucan oligosaccharides

The TXG and BCXG digests were analyzed by HILIC-ESI-CID-MS/MS² on a Vanquish UHPLC system (Thermo Scientific, San Jose, CA, USA) coupled to an LTQ Velos Pro mass spectrometer (Thermo Scientific). The UHPLC settings, column, mobile phases and elution program [29], and the MS (negative ion mode) settings [30] have been described previously. MS² was performed under dependent scan mode.

Quantification and statistical analysis

All enzymatic digestions were performed in technical duplicates and standard deviations were calculated for temperature and pH profiles.

Results and discussion

Discovery of new fungal GH44 enzymes through genome mining

A large set of fungal genomes (>400) were analyzed for their diversity with respect to the set of genes assigned to CAZy families in the CAZy database, including both genomes present in the public CAZy database and genomes only available at JGI MycoCosm [31]. This revealed that part of the basidiomycete genomes contained genes that were already assigned to GH44, a family for which, so far, only bacterial enzymes have been characterized [23]. Interestingly, our genome analyses also revealed that fungal GH44 sequences could only be identified in the phylum Basidiomycota.

GH44 sequences were found in only about 28% of the basidiomycete genomes according to JGI MycoCosm (situation January 22, 2021) (Suppl. Table 1). This narrow distribution among

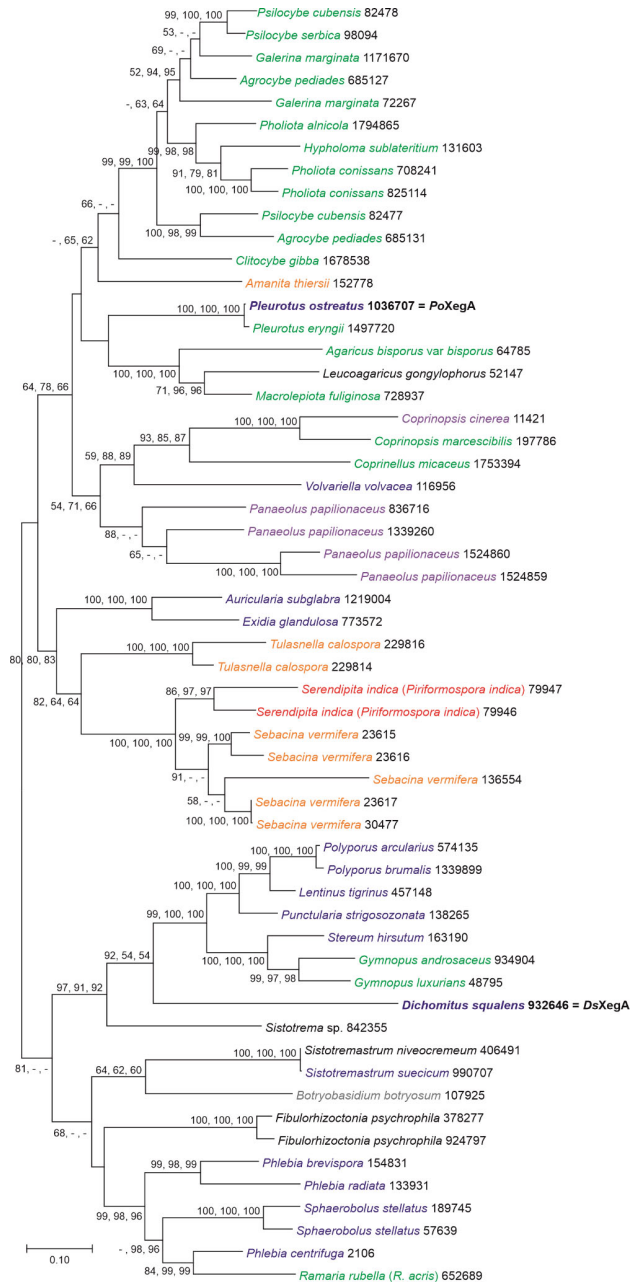


Figure 1 Phylogenetic analysis of GH44 candidates from selected basidiomycete fungi. The figure is a representative Maximum Likelihood tree (500 bootstraps) of a MAFFT alignment of the amino acid sequences of the selected genes. JGI protein IDs are indicated behind the species. Lifestyles of the species are color coded: blue = white rot, grey = grey rot, green = litter decomposer, orange = mycorrhizae, purple = coprophilous, red = endophyte, black = unknown/not clear. Bootstrap values are indicated at the nodes based on Maximum Likelihood, Minimal Evolution and Neighbor Joining algorithms, respectively. Only bootstrap values that are 50 or higher are displayed. The scale bar corresponds to 0.10 amino acid substitutions per site.

basidiomycetes suggests that the GH44 enzymes are ancient, possibly of bacterial origin, and that they were lost during evolution for most of the fungal species, including ascomycetes. Alternatively, a basidiomycete ancestor may have acquired this enzyme by horizontal gene transfer from bacteria after the split from ascomycetes.

Labelling the phylogenetic tree of GH44 candidates with the fungal lifestyles (e.g., white rot, grey rot, litter decomposer, mycorrhizae, coprophilous, and endophyte) demonstrated that GH44 enzymes do not appear associated with a particular lifestyle (**Figure 1**). Some species contain multiple GH44-encoding genes resulting from recent gene duplications, suggesting a more important role for these enzymes in their habitat. One of these fungi is the mycorrhizal fungus *Sebacina vermifera* whose genome encodes five GH44 candidates. It is tempting to speculate that the putative GH44 enzymes may be involved in partial degradation of plant root cell walls that are enriched in xyloglucan, to enable the establishment of symbiosis between fungus and plant roots. Xyloglucan is also released from plant roots [32] and could therefore be a carbon source for fungi that live in close association with roots (e.g., mycorrhizae) or inhabit soil, such as litter-decomposing basidiomycetes, explaining the presence of GH44 enzymes in these fungi (**Figure 1**).

Furthermore, our analysis of the evolutionary relationships between the GH44 candidates showed that those from the two wood degrading white rot fungi, *Dichomitus squalens* XegA and *Pleurotus ostreatus* XegA, were present in distant clades in the phylogenetic tree (**Figure 1**). *In situ* expression and production of the *D. squalens* and *P. ostreatus* GH44 genes and proteins have been shown in previously published transcriptome and exoproteome datasets [33-35]. In short, *D. squalens* CBS464.89 expressed *xegA* and produced the corresponding protein during growth on spruce and birch wood [33, 36], whereas *P. ostreatus* PC9 expressed the XegA encoding gene on rice straw [35], and in the $\Delta pex1$ and $\Delta gat1$ mutants on beechwood sawdust [34], again supporting a role for these genes in plant biomass degradation. Therefore, these two candidates were selected for further characterization to demonstrate the function of basidiomycete GH44 enzymes.

The selected basidiomycete GH44 candidates are xyloglucan-specific XEGs

D. squalens xegA and *P. ostreatus xegA* were heterologously expressed in *Pichia pastoris*, and their recombinant enzymes were produced as C-terminal His-tag fusion proteins and purified from the culture supernatants using immobilized metal affinity chromatography via hexahistidine-tag. The apparent mass of purified enzymes was assessed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and summarized in **Table 1**. The hydrolytic activity of the purified GH44 enzymes toward four types of polysaccharides, including tamarind xyloglucan (TXG), Avicel®PH-101, galactomannan and beechwood xylan, was validated by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Both enzymes were only active toward xyloglucan and not on other tested substrates, which suggested that these enzymes were specific XEGs (further referred to as *DsXegA* and *PoXegA*). The xyloglucan-specific activity of these fungal GH44 XEGs was also observed for fungal XEGs from GH5, GH12 and GH74 [16, 20, 37-41]. In contrast, bacterial GH44 XEGs have been reported to exhibit a broad substrate specificity and, in addition to xyloglucan, other polysaccharides can be hydrolyzed, including barley β -glucan, birchwood xylan, wheat arabinoxylin, carob galactomannan, and synthetic soluble cellulose derivatives [23].

DsXegA and *PoXegA* showed optimum temperature at 50°C (**Figure 2A**), which was slightly lower than that of the reported fungal XEGs of families GH12 and GH74, which typically have optimal activity in a range of 55–65°C [15, 20, 39, 41, 42], but higher than the optimum temperature of *A. oryzae* GH5 XEG, i.e., 45°C [13]. Both XEGs were stable up to 40°C, of which *PoXegA* showed better stability than *DsXegA* at lower temperature (20°C). For pH optimum, *DsXegA* and *PoXegA* showed the highest activity toward TXG at pH 5.0 and 6.0, respectively (**Figure 2B, 2C**). These values are comparable with those of the reported fungal GH5, GH12 and GH74 XEGs that generally have optimal activity in an acidic pH range (3.5–6.0) [16, 20, 37, 39, 40, 42]. When examined for pH stability, *DsXegA* retained more than 75% residual activity after 1 hour of incubation at pH 5.0–7.0,

Table 1 Physical and biochemical properties of purified XGEs used in this study.

GH Family	Fungal Species	Accession Number	Gene Name	Enzyme Name	Calculated Mass (kDa)	Apparent Mass (kDa)		pH optimum	Temperature Optimum (°C)	Positive Substrate ^a	References
						(-) Endo H	(+) Endo H				
44	<i>Dichomitus squalens</i>	JGI Dicsqu464_1 932646	xegA	DsXegA	78	115	90	5.0	50	TXG	This study
44	<i>Pleurotus ostreatus</i>	JGI PleosPC15_2 1036707	xegA	PoXegA	80	90	80	6.0	50	TXG	This study
12	<i>Aspergillus nidulans</i>	XP_658056.1, AN0452	xegA	AnXegA	24	25	25	6.5	47	TXG	[36]
74	<i>Aspergillus nidulans</i>	XP_659146.1, AN1542	xegB	AnXegB	85	100	90	3.0	42	TXG, TXG oligosaccharides	[36, 40]

^a, TXG is tamarind xyloglucan.

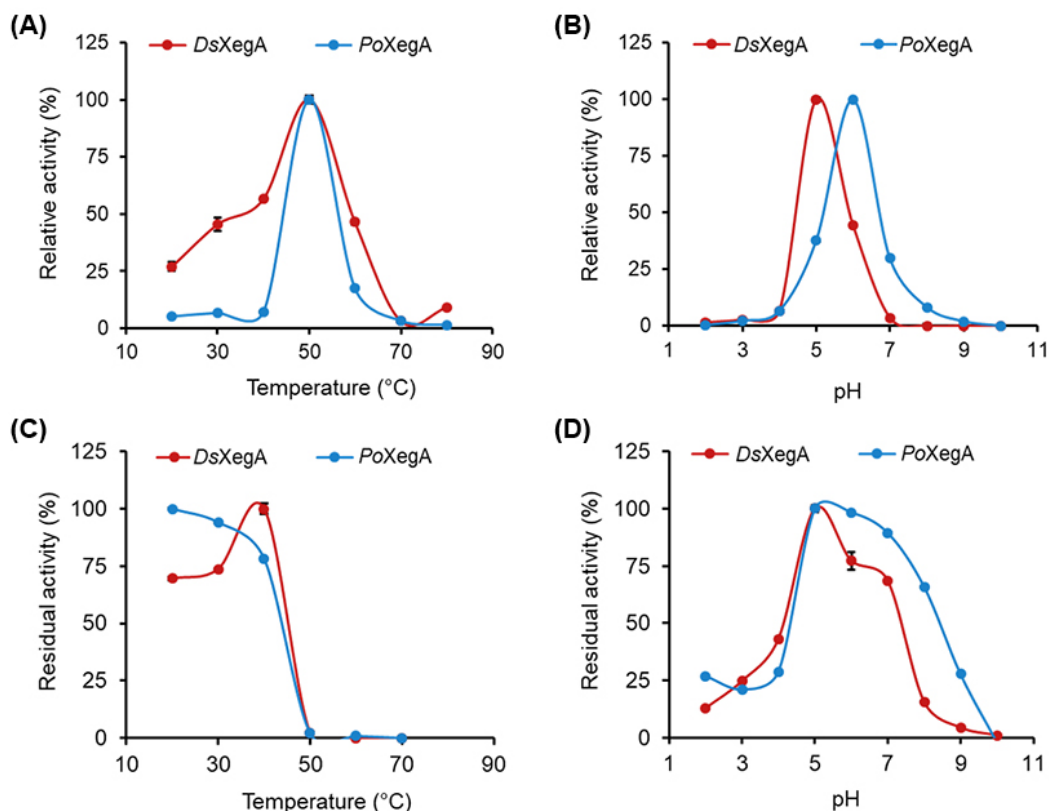


Figure 2 The biochemical properties of *DsXegA* and *PoXegA*. (A) Temperature optimum was measured at 20–80°C in 100 mM phosphate buffer, pH 5.0; (B) pH optimum was measured at 40°C in 100 mM Britton-Robinson buffer, pH 2.0–10.0; (C) Temperature stability was measured at 37°C after pretreatment of enzymes in 100 mM phosphate buffer, pH 5.0, at 20–70°C for 1 h; (D) pH stability was measured at 37°C after pretreatment of enzymes in 100 mM Britton-Robinson buffer, pH 2.0–10.0, at 40°C for 1 h. Error bars depict standard deviation of technical duplicate reactions.

whereas *PoXegA* was stable at the pH range from 5.0 to 8.0 and was more stable than *DsXegA* between pH 6.0 and 8.0 (Figure 2D).

GH44 XEGs generate unique xyloglucan oligosaccharide profiles

To characterize the xyloglucan cleavage patterns by the two GH44 XEGs, the TXG oligosaccharides released by *DsXegA* and *PoXegA* were profiled by HPAEC-PAD (Figure 3A, 3B). For comparison and annotation of peaks, HPAEC profiles of TXG digests generated by *AnXegA* (GH12), by *AnXegB* (GH74), by the previously well characterized *NcLPMO9C* [29] and of the reference TXG oligosaccharides [29] and TXG oligosaccharide standards are added in Figure 3C–3F. *AnXegB* has previously been studied by Bauer, Vasu [43] and *AnXegA* has not been extensively characterized so far [37]. In general, the HPAEC profiles of the *DsXegA* and *PoXegA* digests were similar and only the intensity of several peaks was different. It should be noted that the profiles of the *AnXegA*- and the *AnXegB*-TXG digests were very different, indicating distinct cleavage patterns of TXG by three types of fungal GH44, GH12 and GH74 XEGs used in this study.

Based on the comparison in elution from the standards and reference digests (Figure 3G, 3F), typical XXXG-type oligosaccharides, such as XXXG, XLXG, XXLG and XLLG, were present in

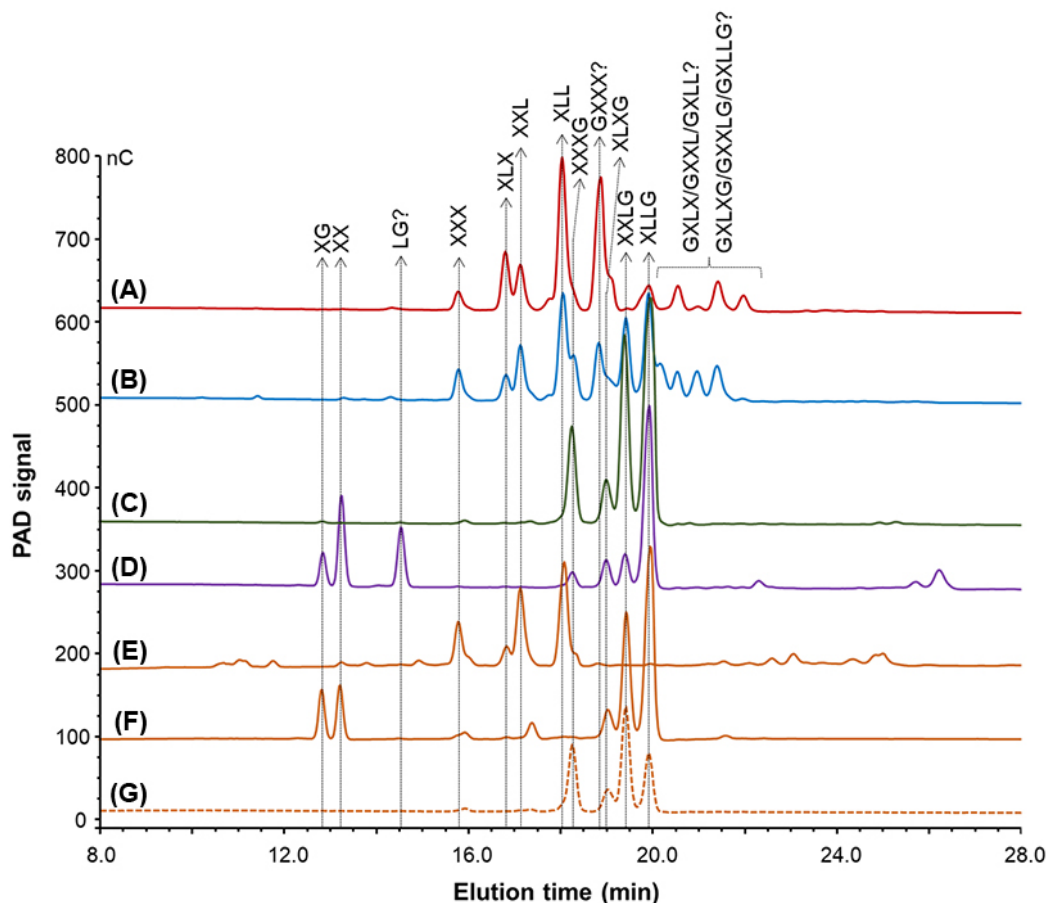


Figure 3 HPAEC elution patterns of tamarind xyloglucan (TXG) oligosaccharides released by end-point (24 h) incubation. TXG oligosaccharides generated by (A) *DsXegA*; (B) *PoXegA*; (C) *AnXegA*; (D) *AnXegB*; (E) *NcLPMO9C* (in the presence of 1 mM ascorbic acid). (F) Reference TXG oligosaccharides; (G) TXG oligosaccharide standards are also shown for annotation of peaks.

all GH44-, GH12- and GH74-TXG digests (**Figure 3A–3D**), although the relative abundance of these compounds differed for each of the digests. In contrast, XXX, XLX, XXL and XLL were only found in the two GH44-digests (**Figure 3C, 3D**). Notably, an unknown peak, eluting between XXXG and XLXG, was identified in both GH44-TXG digests. We propose that this unknown oligomer corresponds to GXXX, based on elution behavior and hydrophilic interaction chromatography coupled with electrospray ionization-collision induced dissociation-mass spectrometry (HILIC-ESI-CID-MS/MS²) analysis described in the next section. Further unique for the GH44-TXG digests was an additional set of yet unknown compounds, eluting after XLLG. Again, taking into account their elution together with the HILIC-ESI-CID-MS/MS² data described in the next section, it could be speculated that these unknown compounds corresponded to GXXXG, GXLX(G), GXXL(G) and GXLL(G).

Furthermore, in the *AnXegB*-TXG digest, XG and XX were identified based on comparison with the reference TXG oligosaccharides (**Figure 3**), suggesting that *AnXegB* was able to cleave XXXG into XX+XG. By analogy with this observation, it could be concluded that XXLG was cleaved into XX and LG, substantiated by the relative low abundance of XXLG. The compound eluting at around 14.5 min was therefore annotated as LG.

Structural elucidation of xyloglucan oligosaccharide released by GH44 XEGs

To further elucidate the structures of xyloglucan oligosaccharides generated by the two GH44 XEGs, *DsXegA*- and *PoXegA*-TXG digests were analyzed by negative ion mode HILIC-ESI-CID-MS/MS². The TXG oligosaccharide standards were also analyzed (**Suppl. Fig. 1**) for the identification and annotation of (unknown) structures. Similar to the HPAEC results, *DsXegA*- and *PoXegA*-TXG digests showed comparable profiles in their base-peak ion chromatograms and only the relative abundance of each product varied (**Suppl. Fig. 1**). Owing to separation of α/β -anomers of TXG oligosaccharides in HILIC, each TXG oligosaccharide split into two distinct peaks. Nevertheless, this α/β -anomer separation did not influence the MS/MS² elucidation of the TXG oligosaccharide structures. MS² fragments of TXG oligosaccharides were annotated following the principle that C-/Z-type, D-type (derived from a double C-/Z-type cleavage of three linked sugar residues), and A-type fragments were generated predominately in negative collision-induced dissociation (CID) ion mode [29, 30, 44, 45].

In both GH44-TXG digests, regular XXXG-type oligosaccharides were analyzed, including XXXG (m/z 1061.6), XLXG (m/z 1223.6), XXLG (m/z 1223.6) and XLLG (m/z 1385.6). Next, “XXX-type” oligosaccharides (XXX, m/z 899.5; XLX and XXL, m/z 1061.6 and XLL m/z 1223.6) and the above suggested “GXXX-type” oligosaccharides (GXXX, m/z 1061.6, GXLL, m/z 1223.6, GXXL, m/z 1223.6 and GXLL, m/z 1385.6) were unambiguously characterized (**Suppl. Fig. 1**). In addition, several “GXXXG-type” oligosaccharides were found in the two GH44-TXG digests, but to a lesser extent. As an example, **Figure 4A and 4B** show the annotation of the MS² fragmentation of XXL and GXXX, respectively. Identification of these two structures was based on the elution time and considering the TXG structure composed of XXXG-type blocks. As shown in **Suppl. Fig. 1**, m/z 1061.6 can represent multiple isomeric structures, not only XXL and GXXX but also XXXG and XLX. Based on the elution time of TXG oligosaccharide standards (31.5–32.1 min), XXXG was first identified (peak number 2 in **Suppl. Fig. 1**) in *DsXegA*- and *PoXegA*-TXG digests, which was also confirmed by the MS² fragments. However, extra MS² fragments m/z 689, 707 and 767 were present in peak number 2 (**Suppl. Fig. 1**), which are slightly different from the MS² fragments of the XXXG standard and indicated the co-elution of other isomers with XXXG. These three fragments represent a C-type glycosidic bond and ^{0,2}A-type cross-ring cleavage to generate a GXX, an XL or an LX fragment (three hexaosyl and two pentaosyl residues) from the non-reducing end. Therefore, these two structures can only be released from XLX or GXXX, while GXXX was identified in the following peak number 4 as described later. To summarize this information, the other isomer in peak number 2 (**Suppl. Fig. 1**) was XLX. Next, we investigated the peak number 3 and 4 also showing oligosaccharides with m/z 1061.6. Interestingly, the oligosaccharide in peak number 3 generated comparable fragments as XXXG standards, but with a more abundant m/z 353 and a lack of ions m/z 899, 839 and 821 (**Figure 4A**). The abundant former fragment indicated the presence of xylosyl-galactosyl residues, and the lack of latter ions often represented the terminal unsubstituted D-glucosyl units at the reducing end side [46, 47]. Taking this information into account, peak number 3 is annotated as XXL structure. Regarding peak number 4 (XLX or GXXX), we detected the diagnostic fragment m/z 413, which resulted from a ^{0,2}A-type cross-ring cleavage on a D-glucosyl residue from either a GX or an L unit at the non-reducing end. As a non-reducing end terminal L unit with parent m/z 1061.6 cannot be present in XLX and GXXX (not even in TXG structure), this diagnostic fragment m/z 413 can only represent a terminal GX structure at the non-reducing end, and therefore, peak number 4 was annotated as GXXX (**Figure 4B**).

This distinct presence of identified XXXG-type, XXX-type and GXXX(G)-type oligosaccharides suggests that the two GH44 XEGs cleaved at both sides of unsubstituted D-glucosyl units in TXG. To further investigate whether the two GH44 XEGs also cleaved next to unsubstituted D-glucosyl units neighbored by more heavily branched F units, we searched for the m/z of 1369.6–1369.8 in the black currant xyloglucan (BCXG) digests. The m/z 1369 represents oligosaccharides having five hexaosyl, three pentaosyl, and one fucosyl residues, which can be XXFG, XFGX, FGXX, GXXF, and XLF. XXFG was first identified based on the diagnostic fragments C₃ (m/z 605) and ^{0,2}A₃ (–H₂O) (m/z 1147 and 1129), which represented a XX fragment at the non-reducing end side (**Suppl. Fig.**

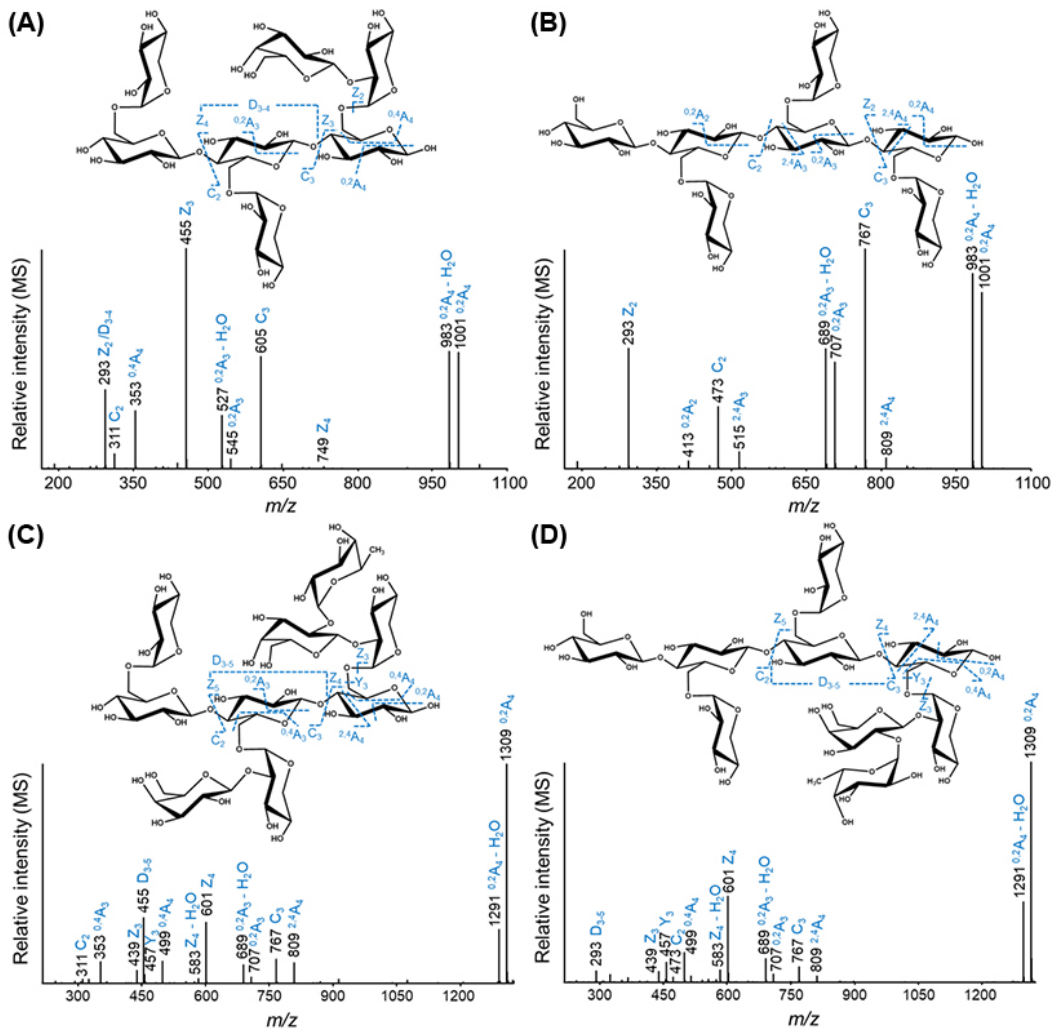


Figure 4 Negative ion mode HILIC-ESI-CID-MS/MS² fragmentation patterns of diagnostic TXG oligosaccharides. (A) Diagnostic XXL (m/z 1061.6); (B) GXXX (m/z 1061.6); (C) XLF (m/z 1369.7); (D) GXXX (m/z 1369.7) both from the end-point *DsXegA*- and *PoXegA*-TXG digests. The fragments are annotated according to the nomenclature proposed by Domon and Costello [43].

2). The non-reducing end side XX fragment can only result from an XXFG unit. Other fragments such as C_4 (m/z 1207), $^{0,2}A_4(-H_2O)$ (m/z 1147 and 1129), D_{2-4} (m/z 601), $^{0,4}A_4$ (m/z 499) and $C_{3\alpha}$ (m/z 457) further confirmed the XXFG structure. In addition to XXFG, XXX-type XLF and GXXX-type GXXX units were released (Figure 4C, 4D). In these MS² spectra (Figure 4C, 4D), fragments $^{2,4}A_4$ (m/z 809), C_3 (m/z 767) and $^{0,2}A_3(-H_2O)$ (m/z 707 and 689) represented fragments composed of three hexaosyl and two pentaosyl residues at the non-reducing end side, which can only be XL and GXX structures in BCXG, and thus XLF and GXXX are the only two possibilities. As shown in the MS² spectrum in Figure 4C, the diagnostic ion m/z 455 can only result from an internal L residue (regarding XLF and GXXX), and thus XLF was identified as such, whereas Figure 4D represented a GXXX unit.

For comparison, *AnXegA*- and *AnXegB*-TXG digests were also analyzed by negative ion mode HILIC-ESI-CID-MS/MS². In brief, *AnXegA* (GH12) only released XXXG-type building blocks

(Suppl. Fig. 3). *AnXegB* (GH74) not only generated XXXG-type building blocks but also released XG (m/z 473.4), XX (m/z 605.4) and LG (m/z 635.4) units explained by the cleavage of XXXG and XXLG (Suppl. Fig. 4).

Distinct xyloglucan cleavage patterns of GH44 XEGs

Based on all identified products in the GH44-TXG and -BCXG digests (series of XXXG-type, XXX-type and GXXX(G)-type oligosaccharides), it was concluded that the GH44 XEGs cleave xyloglucan at both sides of unsubstituted D-glucosyl units and that they are not hindered by substitution of neighboring units (Figure 5A). Different from the GH44 XEGs, the GH12 XEG used in this study (*AnXegA*) only cleaved within xyloglucan at the reducing end side of unsubstituted D-glucosyl units (Figure 5B), which is comparable with other fungal GH12 XEGs [15-17]. The GH74 enzyme (*AnXegB*) cleaved xyloglucan similar to the GH12 enzyme, but with additional cleavages within XXXG and XXLG building blocks to release XX, XG and LG. However, GH74 did not cleave at the reducing end side of L units in XLXG and XXLG (Figure 5C). Similar xyloglucan cleavage patterns

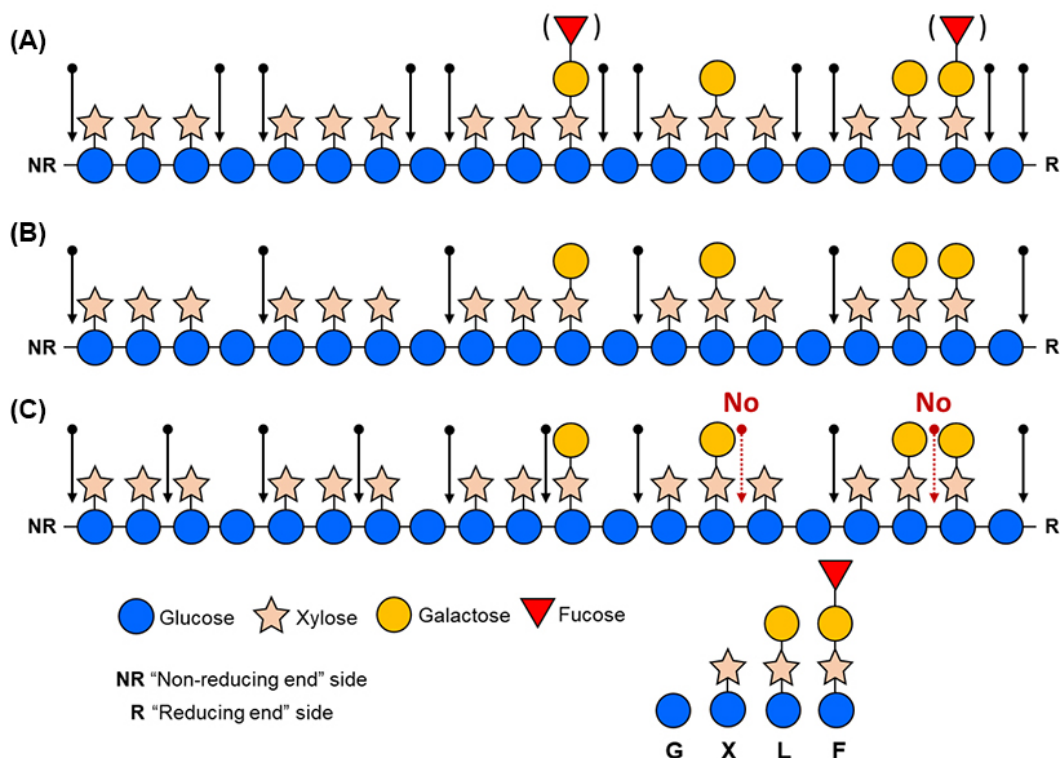


Figure 5 Schematic representation of xyloglucan cleavage patterns by two GH44s (*DsXegA* and *PoXegA*), one GH12 (*AnXegA*), and one GH74 (*AnXegB*). (A) GH44 XEGs cleaved xyloglucan at both sides of unsubstituted D-glucosyl units to release “XXXG-type”, “GXXX(G)-type” and “XXX-type” blocks. The cleavage by GH44 XEGs was not hindered by the substitution of neighboring units. (B) GH12 XEG cleaved in xyloglucan at the “reducing end” side of unsubstituted glucosyl units to release “XXXG-type” blocks. (C) GH74 XEG cleaved xyloglucan at the reducing end side of unsubstituted glucosyl units to release “XXXG-type” blocks and also cleaved within XXXG and XXLG building blocks to release XX+XG and XX+LG. However, GH74 XEG did not cleave at the reducing end side of L units in XLXG and XXLG.

have also been found for other fungal GH74 XEGs [19, 20]. Slightly different from GH74 XEGs, although not studied in this work, the one characterized fungal GH5 XEG can also cleave next to L units to release even G, X, XG, LG, XL/LX, XXG, XXX and LLG units suggested by authors [13]. So far, no fungal GH44 XEG has been reported. For a limited number of bacterial GH44, XEG degrading activity has been reported, but the extent of product profile identification differs and does not match the detail reached in our research. For example, for one bacterial GH44 XEG (*Pp*XG44), it has been shown that the compound XXXGXXXG could be hydrolyzed into XXX and GXXXG [23]. Similarly, for the bacterial Cel44O, identified as an endocellulase, xyloglucan degrading activity and cleavage between XXX and GXXX have been suggested [48].

In conclusion, detailed genome analysis led to the discovery of fungal XEGs in family GH44. Our study provided a comprehensive characterization of two of these fungal XEGs and compared their xyloglucan cleavage patterns with that of fungal XEGs from other GH families. These basidiomycete XEGs are xyloglucan-specific but exhibited distinct xyloglucan cleavage patterns compared with fungal XEGs from families GH12 and GH74. Our study also gave insights on phylogenetic diversity of fungal GH44 and understanding of their roles in biomass (xyloglucan) conversion and valorization.

Author contributions

PS: Investigation, Data curation, Formal analysis, Methodology, Software, Writing – original draft. XL: Investigation, Formal analysis, Validation, Visualization, Writing – original draft. AD: Investigation, Data curation, Formal analysis, Methodology, Validation, Supervision, Writing – original draft, Writing – review & editing. BH: Data curation, Formal analysis, Writing – review & editing. RPDV: Methodology, Supervision, Writing – review & editing, Conceptualization. MAK: Methodology, Supervision, Writing – original draft, Writing – review & editing, Conceptualization. MRM: Investigation, Methodology, Writing – original draft, Writing – review & editing, Conceptualization.

Acknowledgements

BSc Alisa Shubina is acknowledged for her skillful technical assistance. This work was supported by the China Scholarship Council (grant no. 201803250066 to XL) and the Academy of Finland (grant no. 308284 to MRM).

References

1. Scheller, H.V. and P. Ulvskov, *Hemicelluloses*. **Annual Review of Plant Biology**, 2010. 61: p. 263-289.
2. Park, Y.B. and D.J. Cosgrove, *Xyloglucan and its interactions with other components of the growing cell wall*. **Plant and Cell Physiology**, 2015. 56: p. 180-194.
3. Pauly, M., et al., *Changes in the structure of xyloglucan during cell elongation*. **Planta**, 2001. 212: p. 842-850.
4. Koziol, A., et al., *Evaluation of structure and assembly of xyloglucan from tamarind seed (*Tamarindus indica* L.) with atomic force microscopy*. **Food Biophysics**, 2015. 10: p. 396-402.
5. Benkő, Z., et al., *Evaluation of the role of xyloglucanase in the enzymatic hydrolysis of lignocellulosic substrates*. **Enzyme and Microbial Technology**, 2008. 43: p. 109-114.
6. Kaida, R., et al., *Loosening xyloglucan accelerates the enzymatic degradation of cellulose in wood*. **Molecular Plant**, 2009. 2: p. 904-909.
7. Chen, H., et al., *Possible beneficial effects of xyloglucan from its degradation by gut microbiota*. **Trends in Food Science & Technology**, 2020. 97: p. 65-75.
8. Dutta, P., S. Giri, and T.K. Giri, *Xyloglucan as green renewable biopolymer used in drug delivery and tissue engineering*. **International Journal of Biological Macromolecules**, 2020. 160: p. 55-68.
9. Mishra, A. and A.V. Malhotra, *Tamarind xyloglucan: a polysaccharide with versatile application poten-*

- tial. Journal of Materials Chemistry*, 2009. 19: p. 8528-8536.
10. Vincken, J.P., et al., *Two general branching patterns of xyloglucan, XXXG and XXGG. Plant Physiology*, 1997. 114(1): p. 9-13.
 11. Fry, S.C., et al., *An unambiguous nomenclature for xyloglucan-derived oligosaccharides. Physiologia Plantarum*, 1993. 89: p. 1-3.
 12. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Research*, 2014. 42: p. D490-5.
 13. Matsuzawa, T., et al., *Identification and characterization of two xyloglucan-specific endo-1, 4-glucanases in Aspergillus oryzae. Applied Microbiology and Biotechnology*, 2020. 104(20): p. 8761-8773.
 14. Attia, M.A. and H. Brumer, *Recent structural insights into the enzymology of the ubiquitous plant cell wall glycan xyloglucan. Current Opinion in Structural Biology*, 2016. 40: p. 43-53.
 15. Damásio, A.R.L., et al., *Functional characterization and oligomerization of a recombinant xyloglucan-specific endo-beta-1,4-glucanase (GH12) from Aspergillus niveus. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 2012. 1824: p. 461-467.
 16. Master, E.R., et al., *A xyloglucan-specific family 12 glycosyl hydrolase from Aspergillus niger: recombinant expression, purification and characterization. Biochemical Journal*, 2008. 411: p. 161-170.
 17. Vitcosque, G.L., et al., *The functional properties of a xyloglucanase (GH12) of Aspergillus terreus expressed in Aspergillus nidulans may increase performance of biomass degradation. Applied Microbiology and Biotechnology*, 2016. 100: p. 9133-9144.
 18. Martinez-Fleites, C., et al., *Crystal structures of Clostridium thermocellum xyloglucanase, XGH74A, reveal the structural basis for xyloglucan recognition and degradation. Journal of Biological Chemistry*, 2006. 281: p. 24922-24933.
 19. Desmet, T., et al., *An investigation of the substrate specificity of the xyloglucanase Cel74A from Hypocrea jecorina. The FEBS Journal*, 2007. 274: p. 356-363.
 20. Yaoi, K. and Y. Mitsuishi, *Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing end-specific cellobiohydrolase. Journal of Biological Chemistry*, 2002. 277: p. 48276-81.
 21. Arnal, G., et al., *Substrate specificity, regiospecificity, and processivity in glycoside hydrolase family 74. Journal of Biological Chemistry*, 2019. 294: p. 13233-13247.
 22. Attia, M., et al., *Functional and structural characterization of a potent GH 74 endo-xyloglucanase from the soil saprophyte Cellvibrio japonicus unravels the first step of xyloglucan degradation. The FEBS Journal*, 2016. 283: p. 1701-1719.
 23. Ariza, A., et al., *Structure and activity of Paenibacillus polymyxa xyloglucanase from glycoside hydrolase family 44. Journal of Biological Chemistry*, 2011. 286: p. 33890-900.
 24. Linares, N.C., et al., *Recombinant production and characterization of six novel GH27 and GH36 α -galactosidases from Penicillium subrubescens and their synergism with a commercial mannanase during the hydrolysis of lignocellulosic biomass. Bioresource Technology*, 2020. 295: p. 122258.
 25. Katoh, K., G. Asimenos, and H. Toh, *Multiple alignment of DNA sequences with MAFFT*, in *Bioinformatics for DNA sequence analysis. Springer*, 2009: p. 39-64.
 26. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution*, 2016. 33: p. 1870-1874.
 27. Li, X., et al., *Functional validation of two fungal subfamilies in carbohydrate esterase family 1 by biochemical characterization of esterases from uncharacterized branches. Frontiers in Bioengineering and Biotechnology*, 2020. 8: p. 694.
 28. Britton, H.T.S. and R.A. Robinson, *CXCVIII.—Universal buffer solutions and the dissociation constant of veronal. Journal of the Chemical Society (Resumed)*, 1931: p. 1456-1462.
 29. Sun, P., et al., *Configuration of active site segments in lytic polysaccharide monoxygenases steers oxidative xyloglucan degradation. Biotechnol Biofuels*, 2020. 13: p. 95.
 30. Sun, P., et al., *Mass spectrometric fragmentation patterns discriminate C1- and C4-oxidised cello-oligosaccharides from their non-oxidised and reduced forms. Carbohydrate Polymers*, 2020. 234: p. 115917.
 31. Grigoriev, I.V., et al., *MycCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Research*, 2014. 42(D1): p. D699-D704.
 32. Galloway, A.F., et al., *Xyloglucan is released by plants and promotes soil particle aggregation. New Phytologist*, 2018. 217: p. 1128-1136.
 33. Daly, P., et al., *Dichomitus squalens partially tailors its molecular responses to the composition of solid*

- wood. *Environmental Microbiology*, 2018. 20: p. 4141-4156.
34. Wu, H., et al., *Transcriptional shifts in delignification-defective mutants of the white-rot fungus Pleurotus ostreatus*. *FEBS Letters*, 2020. 594: p. 3182-3199.
 35. Wu, H., et al., *Comparative transcriptional analyses of Pleurotus ostreatus mutants on beech wood and rice straw shed light on substrate-biased gene regulation*. *Applied Microbiology and Biotechnology*, 2021. 105: p. 1175-1190.
 36. Daly, P., et al., *Mixtures of aromatic compounds induce ligninolytic gene expression in the wood-rotting fungus Dichomitus squalens*. *Journal of Biotechnology*, 2020. 308: p. 35-39.
 37. Grishutin, S.G., et al., *Specific xyloglucanases as a new class of polysaccharide-degrading enzymes*. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 2004. 1674: p. 268-281.
 38. Bauer, S., et al., *Development and application of a suite of polysaccharide-degrading enzymes for analyzing plant cell walls*. *Proceedings of the National Academy of Sciences of the United States of America*, 2006. 103: p. 11417-11422.
 39. Sinitsyna, O., et al., *Isolation and properties of xyloglucanases of Penicillium sp*. *Biochemistry (Moscow)*, 2010. 75: p. 41-49.
 40. Song, S., et al., *Characterization of two novel family 12 xyloglucanases from the thermophilic Rhizomucor miehei*. *Applied Microbiology and Biotechnology*, 2013. 97: p. 10013-10024.
 41. Xian, L., et al., *Identification and characterization of an acidic and acid-stable endoxyloglucanase from Penicillium oxalicum*. *International Journal of Biological Macromolecules*, 2016. 86: p. 512-518.
 42. Hakamada, Y., S. Arata, and S. Ohashi, *Purification and characterization of a xyloglucan-specific glycosyl hydrolase from Aspergillus oryzae RIB40*. *Journal of Applied Glycoscience*, 2011. 58: p. 47-51.
 43. Bauer, S., et al., *Cloning, expression, and characterization of an oligoxyloglucan reducing end-specific xyloglucanobiohydrolase from Aspergillus nidulans*. *Carbohydrate Research*, 2005. 340: p. 2590-2597.
 44. Quéméner, B., et al., *Negative electrospray ionization mass spectrometry: a method for sequencing and determining linkage position in oligosaccharides from branched hemicelluloses*. *Journal of Mass Spectrometry*, 2015. 50: p. 247-64.
 45. Domon, B. and C.E. Costello, *A systematic nomenclature for carbohydrate fragmentations in FAB-MS/MS spectra of glycoconjugates*. *Glycoconjugate Journal*, 1988. 5: p. 397-409.
 46. Quéméner, B., et al., *Negative electrospray ionization mass spectrometry: a method for sequencing and determining linkage position in oligosaccharides from branched hemicelluloses*. *Journal of Mass Spectrometry*, 2015. 50(1): p. 247-264.
 47. Sun, P., et al., *Mass spectrometric fragmentation patterns discriminate C1- and C4-oxidised cello-oligosaccharides from their non-oxidised and reduced forms*. *Carbohydrate Polymers*, 2020. 234: p. 115917.
 48. Ravachol, J., et al., *Mechanisms involved in xyloglucan catabolism by the cellulosome-producing bacterium Ruminiclostridium cellulolyticum*. *Scientific Reports*, 2016. 6(1): p. 1-17.

Supplementary material

Suppl. Table 1 Fungal species included in the phylogenetic analysis, related to Figure 1.

Suppl. Fig. 1 HILIC-ESI-MS (negative ion mode) analysis of tamarind xyloglucan (TXG) oligosaccharide standards, *DsXegA*-TXG and *PoXegA*-TXG digests. Related to **Figure 4**. (A) HILIC-ESI-MS (negative ion mode) base-peak ion chromatogram of TXG oligosaccharide standards (XXXG, XXLG, XLXG and XLLG). (B) HILIC-ESI-MS (negative ion mode) base peak ion chromatogram of *DsXegA*-TXG digest. (C) HILIC-ESI-MS base-peak ion chromatogram of *PoXegA*-TXG digest. (D) Table of identification and annotation of numbered peaks based on the MS2 fragmentation patterns.

Suppl. Fig. 2 Negative ion mode HILIC-ESI-MS extracted ion chromatograms (m/z 1369.6–1369.8) of end-point black currant xyloglucan (BCXG) digests of *DsXegA* (A) and *PoXegA* (B). Related to **Figure 4**. Negative ion mode HILIC-ESI-CID-MS/MS² fragmentation pattern of XXFG (m/z 1369.7) (C).

Suppl. Fig. 3 HILIC-ESI-MS (negative ion mode) analysis of *AnXegA*-tamarind xyloglucan (TXG) digest. Related to **Figure 4**. (A) HILIC-ESI-MS (negative ion mode) base-peak ion chromatogram of *AnXegA*-tamarind xyloglucan (TXG) digest. (B) Table of identification and annotation of numbered peaks based on the MS² fragmentation patterns.

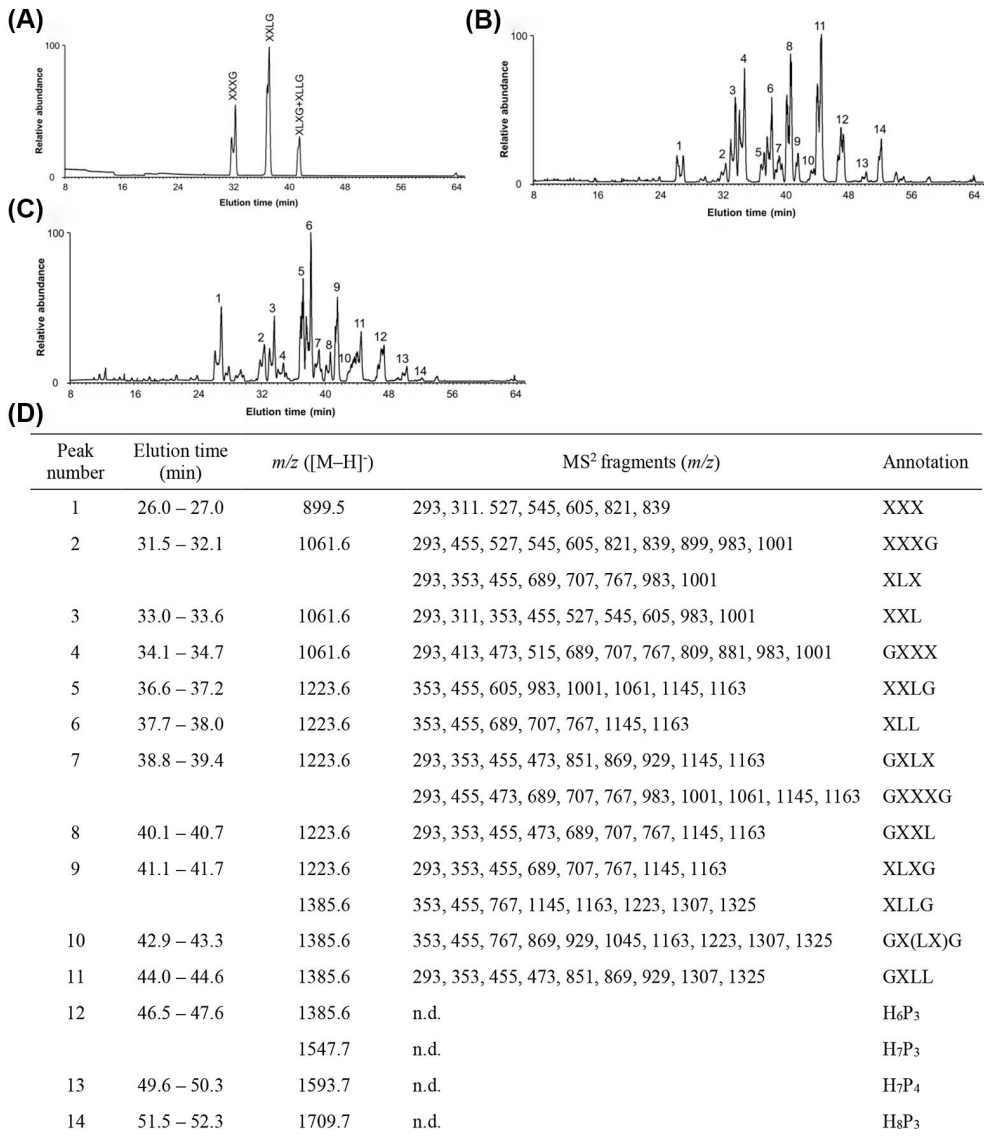
Suppl. Fig. 4 HILIC-ESI-MS (negative ion mode) analysis of *AnXegB*-tamarind xyloglucan (TXG) digest. Related to **Figure 4**. (A) HILIC-ESI-MS (negative ion mode) base-peak ion chromatogram of *AnXegB*-tamarind xyloglucan (TXG) digest. (B) Table of identification and annotation of numbered peaks based on the MS² fragmentation patterns.

Suppl. Table 1 Fungal species included in the phylogenetic analysis.

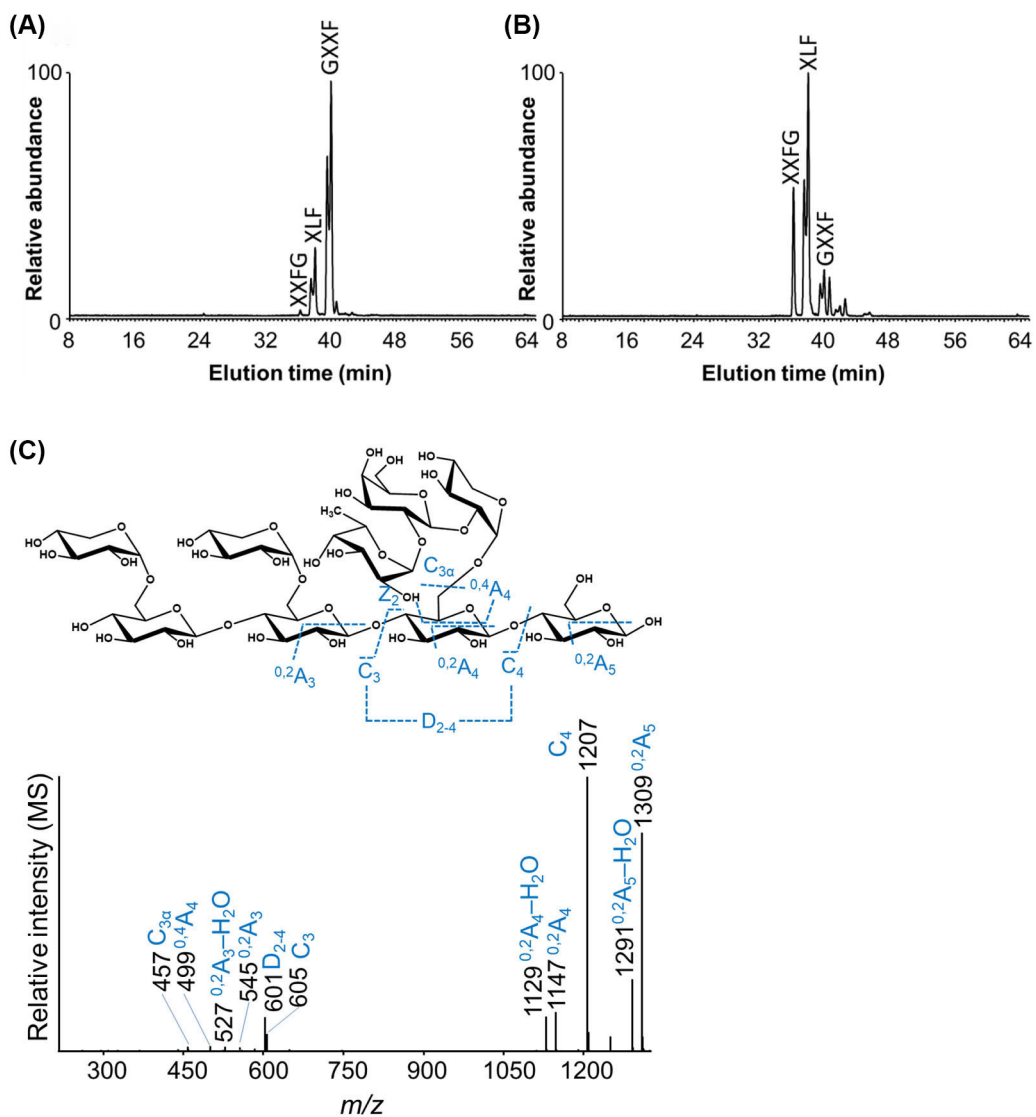
Fungal species and strain	Class	Order	Family	Whole genome sequence at JGI MycoCosm portal
<i>Agaricus bisporus</i> var <i>bisporus</i> (H97)	Agaricomycetes	Agaricales	Agaricaceae	https://mycocosm.jgi.doe.gov/Agabi_varbisH97_2/Agabi_varbisH97_2.home.html
<i>Agrocybe pedatis</i> AH 40210	Agaricomycetes	Agaricales	Bolbitiaceae	https://mycocosm.jgi.doe.gov/Agaped1/Agaped1.home.html
<i>Amanita thiersii</i> Skay4041	Agaricomycetes	Agaricales	Amanitaceae	https://mycocosm.jgi.doe.gov/Amath1/Amath1.home.html
<i>Auricularia subglabra</i>	Agaricomycetes	Auriculariales	Auriculariaceae	https://mycocosm.jgi.doe.gov/Aurde3_1/Aurde3_1.home.html
<i>Botryobasidium botryosum</i>	Agaricomycetes	Cantharellales	Botryobasidiaceae	https://mycocosm.jgi.doe.gov/Boibo1/Boibo1.home.html
<i>Clitocybe gibba</i> UFM A808	Agaricomycetes	Agaricales	Tricholomataceae	https://mycocosm.jgi.doe.gov/Cligib1/Cligib1.home.html
<i>Coprinopsis cinerea</i>	Agaricomycetes	Agaricales	Psathyrellaceae	https://mycocosm.jgi.doe.gov/Copci1/Copci1.home.html
<i>Coprinellus micaceus</i> FP101781	Agaricomycetes	Agaricales	Psathyrellaceae	https://mycocosm.jgi.doe.gov/Copmic2/Copmic2.home.html
<i>Coprinopsis marcescibilis</i> CBS121175	Agaricomycetes	Agaricales	Psathyrellaceae	https://mycocosm.jgi.doe.gov/Copmar1/Copmar1.home.html
<i>Dichomitus squalens</i> CBS464.89	Agaricomycetes	Polyporales	Polyporaceae	https://mycocosm.jgi.doe.gov/Dicsqu464_1/Dicsqu464_1.home.html
<i>Exidia glandulosa</i>	Agaricomycetes	Auriculariales	Auriculariaceae	https://mycocosm.jgi.doe.gov/Exig1/Exig1.home.html
<i>Fibulorhizoctonia psychrophila</i> CBS 109695	Agaricomycetes	Atheliales	Atheliaceae	https://mycocosm.jgi.doe.gov/Fibsp1/Fibsp1.home.html
<i>Galerina marginata</i>	Agaricomycetes	Agaricales	Strophariaceae	https://mycocosm.jgi.doe.gov/Galma1/Galma1.home.html
<i>Gymnopus androsaceus</i> JB14	Agaricomycetes	Agaricales	Omphalotaceae	https://mycocosm.jgi.doe.gov/Gyman1/Gyman1.home.html
<i>Gymnopus luxurians</i>	Agaricomycetes	Agaricales	Omphalotaceae	https://mycocosm.jgi.doe.gov/Gymlu1/Gymlu1.home.html
<i>Hypholoma sublateralium</i>	Agaricomycetes	Agaricales	Omphalotaceae	https://mycocosm.jgi.doe.gov/Hypsu1/Hypsu1.home.html
<i>Lentinus tigrinus</i> ALCF2SS1-6	Agaricomycetes	Polyporales	Polyporaceae	https://mycocosm.jgi.doe.gov/Sisbr1/Sisbr1.home.html
<i>Leucoagaricus gongylophorus</i> Ac12	Agaricomycetes	Agaricales	Agaricaceae	https://mycocosm.jgi.doe.gov/Leugo1_1/Leugo1_1.home.html
<i>Macroleptota fuliginosa</i> MF-1S2	Agaricomycetes	Agaricales	Agaricaceae	https://mycocosm.jgi.doe.gov/Mactu1/Mactu1.home.html
<i>Phlebia brevispora</i> HHB-7030 SS6	Agaricomycetes	Polyporales	Meruliaceae	https://mycocosm.jgi.doe.gov/Phlbr1/Phlbr1.home.html
<i>Phlebia centrifuga</i> FBCC195	Agaricomycetes	Polyporales	Meruliaceae	https://mycocosm.jgi.doe.gov/Phlce1/Phlce1.home.html
<i>Phlebia radiata</i> Fr. (isolate 79, FBCC0043)	Agaricomycetes	Polyporales	Meruliaceae	https://mycocosm.jgi.doe.gov/Phlrad1/Phlrad1.home.html
<i>Pholiotia alnicola</i> AH 47727	Agaricomycetes	Agaricales	Strophariaceae	https://mycocosm.jgi.doe.gov/Phoal1/Phoal1.home.html
<i>Pholiotia conissans</i> CIRM-BRFM 674	Agaricomycetes	Agaricales	Strophariaceae	https://mycocosm.jgi.doe.gov/Phocon1/Phocon1.home.html
<i>Panaeolus papilionaceus</i> CIRM-BRFM 715	Agaricomycetes	Agaricales	Bolbitiaceae	https://mycocosm.jgi.doe.gov/Panpap1/Panpap1.home.html
<i>Pleurotus ostreatus</i> PC15	Agaricomycetes	Agaricales	Pleurotaceae	https://mycocosm.jgi.doe.gov/PleosPC15_2/PleosPC15_2.home.html

Suppl. Table 1 Fungal species included in the phylogenetic analysis (continued).

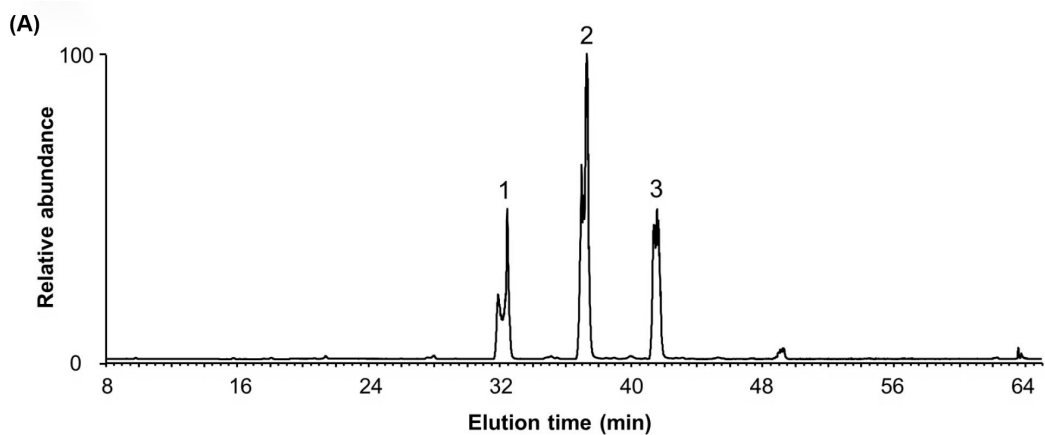
Fungal species and strain	Class	Order	Family	Whole genome sequence at JGI MycoCosm portal
<i>Pleurotus eryngii</i> ATCC 90797	Agaricomycetes	Agaricales	Pleurotaceae	https://mycocosm.jgi.doe.gov/Pleery1/Pleery1.home.html
<i>Polyporus arcularius</i>	Agaricomycetes	Polyporales	Polyporaceae	https://mycocosm.jgi.doe.gov/Polari1/Polari1.home.html
<i>Polyporus brumalis</i> BRFM 1820	Agaricomycetes	Polyporales	Polyporaceae	https://mycocosm.jgi.doe.gov/Polbr1/Polbr1.home.html
<i>Psilocybe cubensis</i>	Agaricomycetes	Agaricales	Hymenogastraceae	https://mycocosm.jgi.doe.gov/Psicub1_1/Psicub1_1.home.html
<i>Psilocybe serbica</i>	Agaricomycetes	Agaricales	Strophariaceae	https://mycocosm.jgi.doe.gov/Psiser1/Psiser1.home.html
<i>Punctularia strigosozonata</i>	Agaricomycetes	Corticiales	Punctulariaceae	https://mycocosm.jgi.doe.gov/Punst1/Punst1.home.html
<i>Ramaria rubella</i> (R. acris) UT-36052-T	Agaricomycetes	Gomphales	Gomphaceae	https://mycocosm.jgi.doe.gov/Ramaci1/Ramaci1.home.html
<i>Sebacina vermifera</i> MAFF 305830	Agaricomycetes	Sebacinales	Serendipitaceae	https://mycocosm.jgi.doe.gov/Sebve1/Sebve1.home.html
<i>Serendipita indica</i> (<i>Piriformospora indica</i>) DSM 11827 from MPI	Agaricomycetes	Sebacinales	Serendipitaceae	https://mycocosm.jgi.doe.gov/Pirin1/Pirin1.home.html
<i>Sistotrema</i> sp. PMI 390	Agaricomycetes	Corticiales	Corticaceae	https://mycocosm.jgi.doe.gov/ClapMI390/ClapMI390.home.html
<i>Sistotremastrum niveocerameum</i> HHB9708 ss-1 1.0	Agaricomycetes	Trechisporales	Hydnodontaceae	https://mycocosm.jgi.doe.gov/Sisni1/Sisni1.home.html
<i>Sistotremastrum suecicum</i>	Agaricomycetes	Trechisporales	Hydnodontaceae	https://mycocosm.jgi.doe.gov/Sissu1/Sissu1.home.html
<i>Sphaerobolus stellatus</i>	Agaricomycetes	Gastrales	Geastraceae	https://mycocosm.jgi.doe.gov/Sphst1/Sphst1.home.html
<i>Stereum hirsutum</i> FP-91666 SS1	Agaricomycetes	Russulales	Stereaceae	https://mycocosm.jgi.doe.gov/Steh1/Steh1.home.html
<i>Tulasnella calospora</i> AL134D	Agaricomycetes	Cantharellales	Tulasnellaceae	https://mycocosm.jgi.doe.gov/Tulca1/Tulca1.home.html
<i>Volvariella volvacea</i> V23	Agaricomycetes	Agaricales	Pluteaceae	https://mycocosm.jgi.doe.gov/Volvo1/Volvo1.home.html



Suppl. Fig. 1 HILIC-ESI-MS (negative ion mode) analysis of tamarind xyloglucan (TXG) oligosaccharide standards, *DsXegA*-TXG and *PoXegA*-TXG digests. Related to **Figure 4**. (A) HILIC-ESI-MS (negative ion mode) base-peak ion chromatogram of TXG oligosaccharide standards (XXXG, XXLG, XLXG and XLLG). (B) HILIC-ESI-MS (negative ion mode) base peak ion chromatogram of *DsXegA*-TXG digest. (C) HILIC-ESI-MS base-peak ion chromatogram of *PoXegA*-TXG digest. (D) Table of identification and annotation of numbered peaks based on the MS² fragmentation patterns.



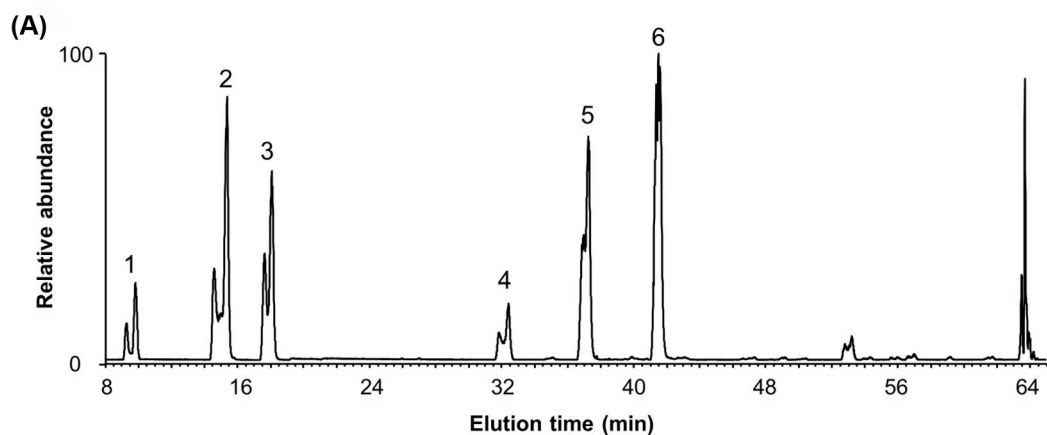
Suppl. Fig. 2 Negative ion mode HILIC-ESI-MS extracted ion chromatograms (m/z 1369.6–1369.8) of end-point black currant xyloglucan (BCXG) digests of *DsXegA* (A) and *PoXegA* (B). Related to Figure 4. Negative ion mode HILIC-ESI-CID-MS/MS² fragmentation pattern of XXFG (m/z 1369.7) (C).



(B)

Peak Number	Elution time (min)	m/z ([M-H] ⁻)	MS ² fragments (m/z)	Annotation
1	31.9 – 32.5	1061.6	293, 455, 527, 545, 605, 821, 839, 899, 983, 1001	XXXG
2	36.6 – 37.2	1223.6	353, 455, 605, 983, 1001, 1061, 1145, 1163	XXLG
3	41.1 – 41.5	1223.6	293, 353, 455, 689, 707, 767, 1145, 1163	XLXG
		1385.6	353, 455, 767, 1145, 1163, 1223, 1307, 1325	XLLG

Suppl. Fig. 3 HILIC-ESI-MS (negative ion mode) analysis of *AnXegA*-tamarind xyloglucan (TXG) digest. Related to Figure 4. (A) HILIC-ESI-MS (negative ion mode) base-peak ion chromatogram of *AnXegA*-tamarind xyloglucan (TXG) digest. (B) Table of identification and annotation of numbered peaks based on the MS² fragmentation patterns.



(B)

Peak Number	Elution time (min)	m/z ([M-H] ⁻)	MS ² fragments (m/z)	Annotation
1	9.2 – 9.9	473.4	131, 149, 161, 221, 293, 311, 395, 413	XG
2	14.6 – 15.3	605.4	149, 191, 293, 311, 527, 545	XX
3	17.6 – 18.3	635.4	161, 179, 293, 311, 353, 413, 455, 473, 557, 575	LG
4	31.9 – 32.5	1061.6	293, 455, 527, 545, 605, 821, 839, 899, 983, 1001	XXXG
5	36.6 – 37.2	1223.6	353, 455, 605, 983, 1001, 1061, 1145, 1163	XXLG
6	41.1 – 41.5	1223.6	293, 353, 455, 689, 707, 767, 1145, 1163	XLXG
		1385.6	353, 455, 767, 1145, 1163, 1223, 1307, 1325	XLLG

Suppl. Fig. 4 HILIC-ESI-MS (negative ion mode) analysis of *AnXegB*-tamarind xyloglucan (TXG) digest. Related to Figure 4. (A) HILIC-ESI-MS (negative ion mode) base-peak ion chromatogram of *AnXegB*-tamarind xyloglucan (TXG) digest. (B) Table of identification and annotation of numbered peaks based on the MS² fragmentation patterns.

Chapter 7

Summary and general discussion

Enzymes are biological catalysts facilitating a wide range of chemical transformations, often with high specificity under very mild reaction conditions [1]. The addition of enzymes in modern industrial processes can significantly reduce energy consumption and waste generation, which has led to an increased use of enzymes as 'green' catalysts, especially in the utilization of lignocellulosic biomass [1]. More specific, by means of an ideal enzyme cocktail, polysaccharides in lignocellulosic biomass can be fully converted into soluble monosaccharides, which can be further used for the manufacture of chemical building blocks or different high-value chemicals [2, 3]. This specific aim has forced the continuous development of novel carbohydrate degrading enzymes with unique properties and the renewal or expansion of existing enzyme cocktails.

Industrial enzyme producers prefer enzymes which are straightforward to produce, hence of microbial origin, especially from fungi and bacteria, while those of plant, animal, and insect origin are less favorable. In particular, fungi are considered to be very good enzyme producers, and fungal based enzymes comprise more than 50% of the total enzyme market [4]. This huge market share is mainly attributed to a few species of *Aspergillus* and *Trichoderma* with desirable industrial characteristics, such as *Aspergillus niger*, *Trichoderma reesei* [5, 6]. These two fungi, in contrast to many others, stand out with respect to a high enzyme protein production level, no toxin production, and a fast growth behavior under standard conditions. To further benefit from the potential of fungi and enzymes present throughout the full fungal kingdom, in recent years, both heterologous expression of functional genes in high-yielding strains using recombinant DNA technology and strain engineering using CRISP/Cas 9 technology has become popular [7-9].

With the high increase in sequenced fungal genomes, the number of genes encoding putative CAZymes has increased exponentially [10, 11]. Only a limited number of the corresponding CAZymes have, however, been biochemically characterized, and to date their functions are assumed and annotated based on the functions of the CAZy (sub)family they belong to or on homology to a characterized enzymes. This approach may lead to misinterpretations, since it cannot be excluded that even within established CAZy families, activities will be found that have so far not been listed.

Pichia pastoris is a popular host for heterologous protein production, as it can secrete high titers of properly folded, post-translationally processed and active recombinant fungal proteins into the culture media [12]. Empirically, proteins secreted by their native fungal host, are usually secreted by *P. pastoris* as well [12]. However, in order to keep a good percentage of characterized reference enzymes versus only annotated ones with putative functions, expression of unknown CAZyme-encoding genes from fungal genomes in *P. pastoris* and comparative characterization of the produced enzymes needs to be extended in the future. This will likely also provide novel enzymes with desirable characteristics for industrial applications.

Xylan is the most abundant representative of plant based hemicellulose, and its complete degradation plays a key role in the overall utilization of hemicellulosic biomass. The main focus of this thesis is to characterize CAZymes encoded by unknown genes involved in xylan degradation in fungal genomes. Using genome mining combined with phylogenetic analysis, we have successfully discovered several novel xylanolytic enzymes including xylan-backbone degrading enzymes (i.e., endoxyxylanase (XLN), xylobiohydrolase (XBH)), and xylan-side chain degrading enzymes (e.g., acetyl xylan esterase (AXE), feruloyl esterase (FAE), α -L-arabinofuranosidase (ABF)). This strategy of expanding the enzyme toolbox was also demonstrated in the discovery of other novel activities, such as endoxyloglucanase (XEG). The enzymes characterized in this project have unique properties and can be used as new tools for biotechnology. This PhD thesis provides insights in the CAZyme expansion in fungal genomes and deepens our understanding of the diversity of CAZy family activities.

In the following sections, the strategy of enzyme discovery and the diversity within CAZy families are discussed separately. The discussion in context of enzyme discovery focusses on encountered considerations important for the discovery strategy of novel enzymes. The new enzymes described in this thesis illustrate the functional diversity within and between CAZy families. However, there is still huge room for better understanding CAZy families, and therefore the discussion of the diversity

of CAZyme families in this section focuses on possible future expansion of research aspects based on existing studies. The eventual rationale of discovering (novel) enzymes with various specific functionalities is to renew or make enzyme cocktails to apply for efficient degradation of plant biomass. In this thesis, I focused on the enzyme discovery and biochemical characterization by using model substrates and purified plant polysaccharides. The actual synergistic degradation of total plant biomass was considered out-of-scope of this thesis, although at the end of this discussion section, I briefly provide some ideas on how to set up experiments to evaluate degradation of plant biomass using the enzymes found in this study.

Considerations for enzyme discovery

In **Chapter 1** of this thesis, we described different strategies for enzyme discovery, amongst which the genome mining strategy has become increasingly popular with the explosion of available fungal genome sequences. The genome mining approach can quickly and easily target candidate enzymes with a putative function from fungal genome data, but characterization of these predicted enzymes still requires extensive laboratory work, including the expression of the genes encoding the predicted enzymes in a suitable host, production and purification of the enzymes, and determination of their biochemical properties. Therefore, it is particularly important to select representative sequences from the huge set of sequences that are available. The use of genome mining in combination with amino acid sequence-based phylogeny provides a perfect starting point for the selection of reference enzymes [13-15]. We validated the feasibility of this approach in the discovery of novel enzymes in several chapters of this thesis (**Chapter 2, 3, 6**). Overall, the discovery of novel enzymes in our study includes mainly the following steps, i.e., the selection of fungal genomes, the collection of selected sequences, the analysis of sequence phylogeny, the choice of potential candidates, and the characterization of target candidates (see **Figure 1**). The characterization of selected candidates will further contribute to the correct annotation of fungal genomes. As a reference for researchers working in related areas, here I discuss several challenges that need to be overcome in order to accelerate the genome mining process and ultimately support the discovery of novel enzymes (see **Figure 1**).

Firstly, the selection of fungal genomes to be included is important (see **Figure 1**). The explosion of high-throughput sequencing approaches and analytical tools has made sequencing of organisms easier and cheaper [16]. Currently, more than 3700 fungal genomes are available in GenBank (<https://www.ncbi.nlm.nih.gov/genome/?term=fungi>). However, the quality of genome sequences has been a concern due to challenges in assembling repeat-rich regions and variable or insufficient sequencing coverage [17]. The use of different methods, e.g., BUSCO [18] with the Fungi OrthoDB v9 database [19] to assess genome quality, and thus select high-quality fungal genomes to mine for predicted CAZyme-encoding genes, is crucial for the successful discovery of novel enzymes. In addition, another consideration is the taxonomic diversity of fungi. For fungi that depend in some way on living or dead plants for their energy supply, their diversity is often closely related to the decomposition process of the carbon-containing plant polymers [20, 21]. In detail, fungi secrete a series of enzymes to degrade the corresponding carbon-containing polymers, releasing the low molecular weight compounds that are able to enter fungal cells as C-source and can be converted into energy and cellular building blocks through specific metabolic pathways. Therefore, the variety and content of CAZymes secreted by fungi are a factor in our assessment of fungal diversity. Overall, this thesis focusses on fungal species that cover a wide part of the fungal kingdom (dikarya), have good genome quality, and are rich in CAZymes.

Secondly, correction of gene models encoding the predicted enzymes mined from fungal genomes is a major requirement (see **Figure 1**). Sequences obtained from different fungal genomes (e.g., from the JGI MycoCosm database, <https://mycocosm.jgi.doe.gov/mycocosm/home>) are used together with characterized enzymes (from the CAZy database, <http://www.cazy.org/>) to construct phylogenetic trees to gain insight into the classification status of sequences without functional description [22, 23]. Thus, reliable phylogenies should be based on correct amino acid sequences [15]. However,

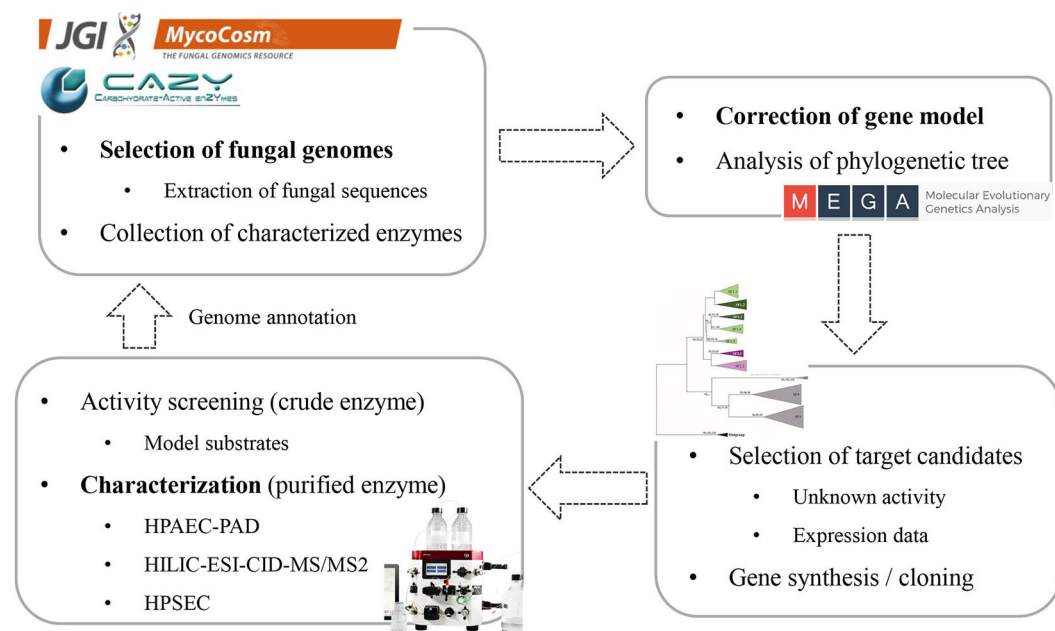


Figure 1 Overall concept of enzyme discovery through the genome mining strategy. Challenging steps are in boldface and discussed in the “Considerations for enzyme discovery process” section.

unfortunately, incorrect gene models are commonly present in fungal genomes even if fungal genomes with good quality are selected. These errors are particularly common in the prediction of the start codon (ATG), stop codon (TAG, TAA, and TGA), or in the intron-exon boundaries [15]. Erroneous gene models result in longer or shorter amino acid sequences thus affecting the overall distribution of sequences in the tree, leading to incorrect analysis in the phylogeny. Therefore, manual curation of gene models is crucial to achieve a correct prediction, which helps to select optimal candidates from phylogenetic tree. As we mentioned above, the errors generally result from the automatic gene calling that follows relatively strict rules about start/stop codon and intron/exon boundaries. There are however many genes that don't strictly follow the consensus. This results in manual curation work ranging from minutes to hours or even days for each gene. Because this is so time consuming, it is often omitted in enzyme mining studies.

Thirdly, characterization of the selected reference enzymes using multiple approaches is necessary (see Figure 1). The selected candidates are produced and characterized in *P. pastoris*. Generally, enzymes with AXE and FAE activity (Chapter 2) can be firstly assayed using methyl umbelliferyl or methyl hydroxycinnamate as substrate. Enzymes with exo-activity (e.g., ABF, Chapter 4) and endo-activity (e.g., XLN, Chapter 5) can be measured using methods of para-nitrophenol (*p*NP) assay (e.g., *p*NP- α -L-arabinofuranoside as substrate in Chapter 4) and 3,5-dinitrosalicylic acid (DNS) reducing sugar assay (e.g., beechwood xylan as substrate in Chapter 5), respectively. These assays can all be performed in microplate reader using 96-well plates. The most obvious advantage of these methods is that they do not require laborious samples treatment, and they can be performed within minutes rather than hours. Also, multiple samples can be measured simultaneously when combined in the same 96-well plate by a microplate reader. However, the main disadvantage of such assays to consider is that all depend on model substrates or model reactions rather than degradation of true plant polysaccharides. In addition, side reactions, for example due to impurities, are hard to pick up with standard assays as side products are usually undetected or even interfering. Nevertheless, when assaying enzymes with purified plant polysaccharides, their (oligosaccharide) product profiles can be complex. To overcome this drawback, additional analytical approaches have

been developed to characterize enzyme products in more detail, such as high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) (used in **Chapter 2, 3, 5, 6**), and hydrophilic interaction chromatography electrospray ionization negative ion mode collision-induced dissociation mass spectrometry (HILIC-ESI-CID-MS/MS²) (used in **Chapter 6**). HPAEC-PAD is a powerful technique to separate and detect complex mixtures of oligosaccharides, and thus can be used to monitor and compare hydrolysis product profiles. However, the eluents used for this technique are not compatible with mass spectrometric (MS) detection, and therefore, the identification of the oligosaccharide structures is hampered. HILIC-ESI-CID-MS/MS², although resulting in a lower chromatographic resolution compared to HPAEC, is compatible with MS⁽ⁿ⁾ detection, and used to determine the structures of generated products in detail. Detailed product characterization data give insight into the polysaccharides cleavage pathways by the (characterized) enzymes, which is highly valuable information for future applications. At the same time, these characterized enzymes facilitate reliable annotation of the genomes of other fungi and further help to rapidly predict the plant biomass degrading capacity of the corresponding species [15].

New insights into the diversity within CAZy families

The characterization of unknown enzymes has provided some insight into the diversity of activities and subfamilies within CAZy families. In short, for CE1 five subfamilies have been listed, amongst which novel enzymes with dual FAE/AXE activity were found in CE1_2 (**Chapter 2**), while GH30 has 11 subfamilies, among which novel enzymes with disaccharide hydrolase activity were found in GH30_5, GH30_7, and GH30_11. Additionally, GH30_11 was proposed as new subfamily in GH30 (**Chapter 3**). *Penicillium subrubescens* contains multiple ABFs and XLNs within the same or different GH families. Functional characterization of these enzymes showed that, again, enzymes from the same GH family had a different substrate specificity (e.g., GH62 arabinoxylan arabinofuranohydrolases from **Chapter 4**) and different detailed product profiles (e.g., GH10 XLNs from **Chapter 5**). Moreover, GH44 has previously been reported to only contain bacterial XEGs, but fungal genome mining combined with experimental characterization showed that this family also contained fungal XEGs (**Chapter 6**). In the remainder part of this section, each studied CAZy family in this thesis will be further discussed in addition to aspects of interest for future research.

As mentioned above, in the CE1 family in **Chapter 2**, we characterized one enzyme with dual FAE/AXE activity (FaeD, **Table 1**) in CE1_2, a subfamily that previously contained only characterized enzymes with FAE activity. Recently, it was found that enzymes with dual FAE/AXE (named FXE) activity were also present in CE1_5 [24]. Similar to CE1_2, CE1_5 was previously predicted to be a subfamily with only FAE activity. Up to this point, it is clear that these two subfamilies are functionally close. However, in the CE1 phylogenetic tree, CE1_2 and CE1_5 are clearly separated (**Fig. 2**), which may reflect some differences in their substrate specificity regarding to the type of the hydroxycinnamic acid (e.g., ferulic, di-ferulic or *p*-coumaric acid) or the type of monosaccharide to which the hydroxycinnamic acid is esterified to.

The presence of hydroxycinnamic acids, including ferulic acid (4-hydroxy-3-methoxy-cinnamic acid), *p*-coumaric acid (4-hydroxy-cinnamic acid), caffeic acid, and sinapic acid, in the cell walls of different plant species has long been known, with ferulic acid and *p*-coumaric acid being common constituents of the cell walls of several plant families [25]. These hydroxycinnamic acids are introduced into plant cell wall polysaccharides through ester linkages between their carboxylic acid group and the primary alcohol on the C-5 of L-arabinosyl substitution of the arabinoxylan backbone, of which ferulic acid can be further covalently linked to the C-9 position of lignin monomers through ether bonds [26-28]. FAE can hydrolyze the ester linkages between hydroxycinnamic acids and plant cell-wall polysaccharides, however, the hydrolysis preference of FAEs or FXEs from CE1_2 and CE1_5 for ferulic acid and *p*-coumaric acid are currently uncertain. In addition, it is noteworthy that ferulic acid was reported to be esterified with different monosaccharides at diverse positions, in addition to the one mentioned above (position C-5 to L-arabinose) [29, 30]. For example, ferulic acid could be linked at position C-2 to L-arabinosyl residues in arabinans, position C-6 to D-galactosyl

Table 1 Overview of the enzymes that have been characterized in this study.

Enzyme activity	Enzyme name	CAZy family	Polysaccharides	Product	Chapter		
Main-chain-degrading enzymes	<i>PsXlnA</i>	GH10	Xylan	Xylo-oligosaccharides; D-xylobiose	5		
	<i>PsXlnB</i>						
	<i>PsXlnC</i>						
	<i>PsXlnD</i>	GH11					
	<i>PsXlnE</i>						
	<i>PsXlnF</i>						
	<i>PsXlnG</i>						
	<i>PsXlnH</i>						
	<i>PsXlnI</i>						
	<i>PsXlnJ</i>						
<i>PsExlA</i>	GH30_7						
Endoxylanase / Xylobiohydrolase (XLN / XBH)		<i>TtExlA</i>					
Xylobiohydrolase (XBH)	<i>TeXbhA</i> <i>TtXbhA</i>						
Endoxyloglucanases (XEG)	<i>DsXegA</i> <i>PoXegA</i>	GH44	Xyloglucan	Xyloglucan-oligosaccharides	6		
Endoglucanase (ENG)	<i>PsEngA</i> <i>AnEngA</i> <i>TaEngA</i> <i>TeEngA</i>	GH30_3	β -(1→6)-glucan	Glucooligosaccharides	3		
Side-chain-degrading enzymes	Acetyl xylan esterase (AXE)	<i>AxeA</i> <i>AxeB</i> <i>Axe1</i>	CE1_1	Xylan	Acetic acid	2	
	Feruloyl esterase / Acetyl xylan esterase (FAE / AXE)	<i>FaeD</i>	CE1_2	Xylan; Pectin	Acetic acid; ferulic acid		
	α -L-Arabinofuranosidase (ABF)	<i>PsAbfA</i> <i>PsAbfC</i>	GH51	Xylan; Pectin; Xyloglucan; Type II arabinogalactan	L-Arabinose	4	
		<i>PsAbfB</i> <i>PsAbfE</i> <i>PsAbfF</i>	GH54				
		Arabinoxylan α -L-arabinofuranohydrolase (AXH)	<i>PsAxA</i> <i>PsAxB</i> <i>PsAxC</i> <i>PsAxD</i>				GH62
		Galactobiohydrolase (GBH)	<i>PsGbhA</i> <i>TvGbhI</i> <i>ShGbhA</i>				GH30_5 GH30_11

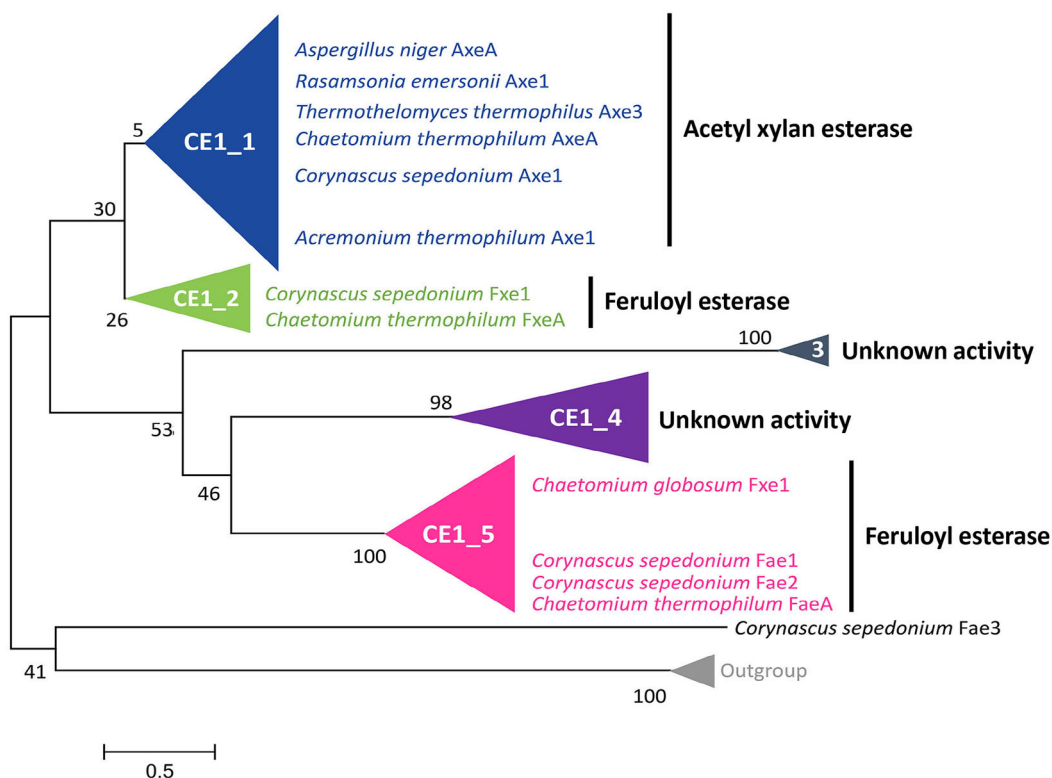


Figure 2 Phylogenetic relationships among the fungal CE1 members (Based on [24]). The evolutionary history was inferred using the Neighbor-Joining method with a bootstrap value of 500. Eight non-CE1 FAEs from SF7 [32] were used as an outgroup.

residues in galactans and position C-4 to D-xylosyl side chains in xyloglucans [29, 30]. The present study in **Chapter 2** does not clarify the specificity of the enzymes from CE1_2 and CE1_5 regarding the position of monosaccharide esterification. In addition, subfamilies with unknown activity, e.g., CE1_3 and CE1_4, would require characterization in order to gain a comprehensive understanding of the CE1 family. Therefore, a larger number of enzymes from CE1 and their specificity for substrates differing in the esterified position of L-arabinose or other sugars need to be investigated in the future. A full understanding of each subfamily in CE1 will help to select optimal candidates for the desired application, such as the bioproduction of high-purity ferulic acid for biosynthesis of vanillin in the flavor market, high-purity *p*-coumaric acid as skin-lightening active ingredient in cosmetics [31].

In the GH30 family, GH30_5 previously contained only characterized enzymes with endogalactanase activity from ascomycetes. The additional characterizations of GH30_5 in **Chapter 3** showed that GH30_5 also included ascomycete enzymes with galactobiohydrolase (GBH) activity, which is similar to the found activity of GH30_11 (**Table 1**). However, GH30 phylogenetic analysis suggested that the fungal GH30_11 were more related to the bacterial GH30_4 than to GH30_5 (**Figure 3**). Intriguingly, β -D-fucosidase activity was previously found in GH30_4. These analyses raised some questions that will need to be clarified in the future. For example, whether GH30_4 and GH30_11 contain other activities we have not yet found, and whether enzymes with the same activity in GH30_5 and GH30_11 behave differently during the hydrolysis of substrates. Therefore, additional candidates need to be selected and characterized in GH30_4 and GH30_11 to explore functional diversity in GH30_4 and GH30_11, while more detailed analyses with regard to substrate specificity and homology modelling need to be done for characterized enzymes in GH30_5 and GH30_11 to understand their potential differences. For the latter, due to the lack of reliable crystal

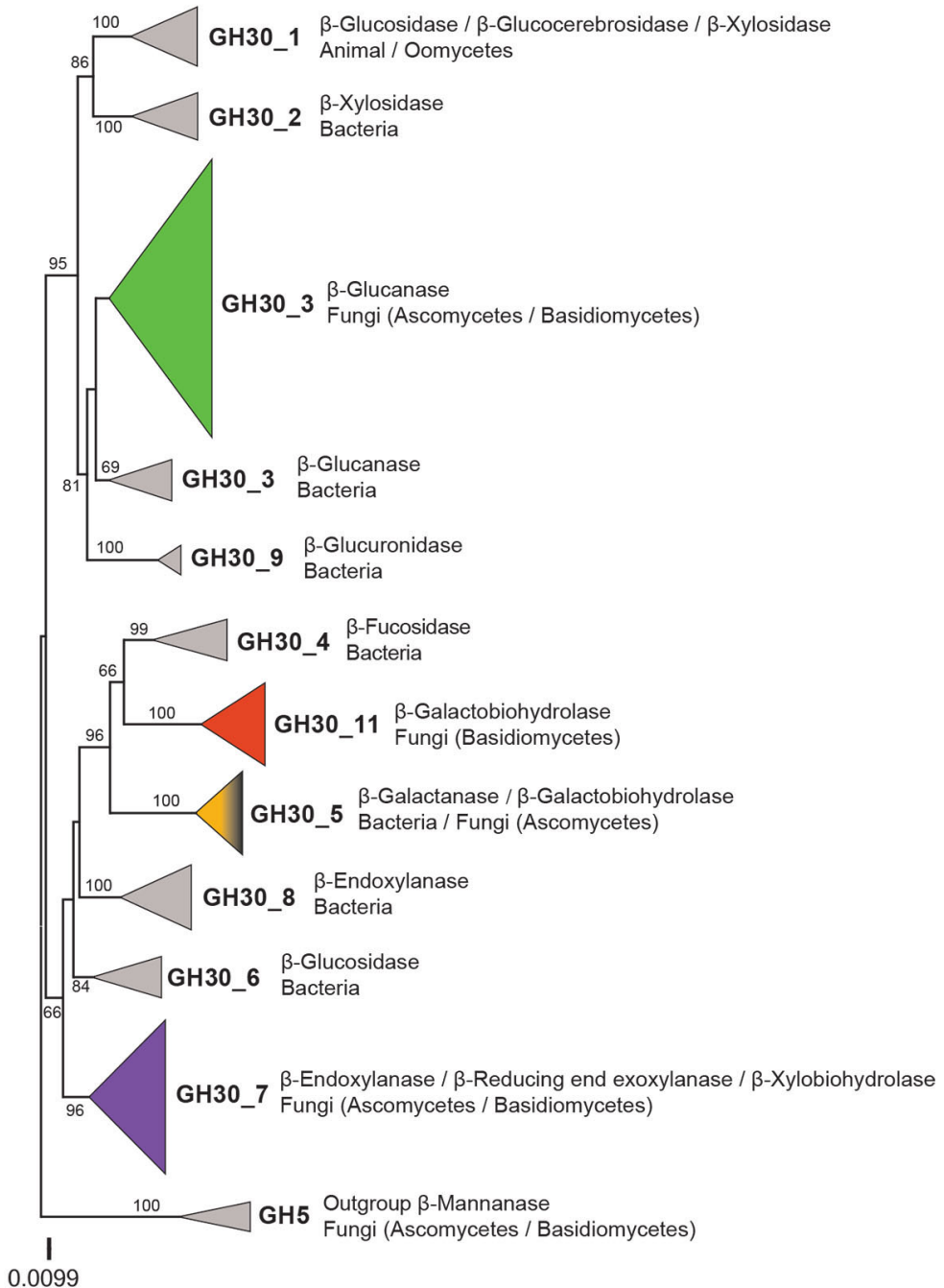


Figure 3 Phylogenetic relationship of GH30 members from fungi, bacteria, and animals based on their amino acid sequences. (Based on [33]). The main subfamilies were collapsed. Statistical support for phylogenetic grouping was estimated by 500 bootstrap re-samplings, only the bootstrap above 50% were shown on the clades. Five mannanases from GH5 were used as an outgroup.

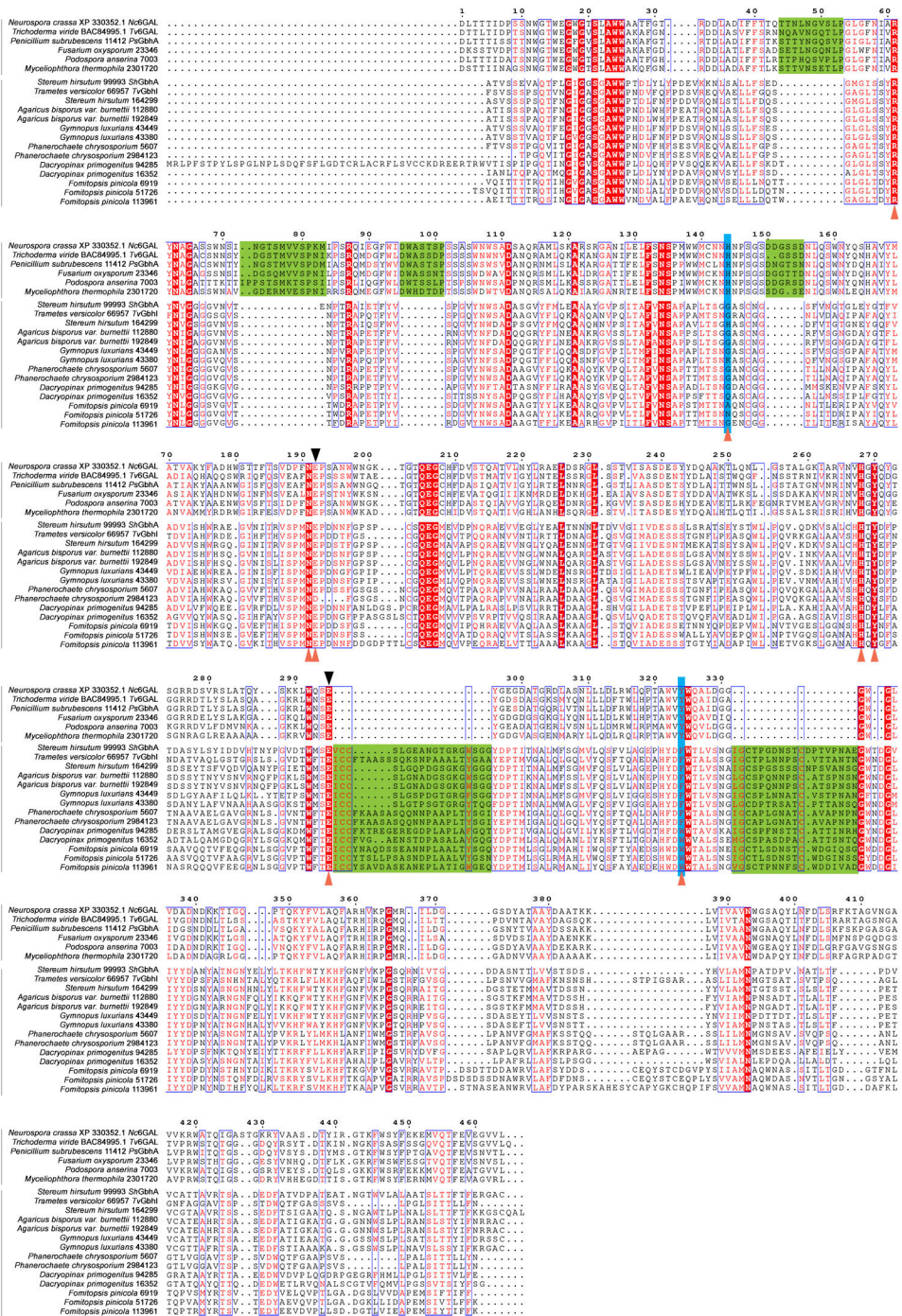


Figure 4 Amino acid sequence alignment of fungal members from GH3_5 and GH3_11 (Sequences from [33]). The putative catalytic residues and conserved amino acids in the catalytic domains are marked by black and orange triangles, respectively (based on [34]). The difference between conserved amino acids of GH3_5 and GH3_11 are highlighted with blue boxes. The different insertions in GH3_5 and GH3_11 are highlighted with green boxes.

structures in both subfamilies, homology modelling analysis has not been feasible so far. However, the sequence alignment revealed some differences among GH30_5 and GH30_11 members (**Figure 4**). For example, the conserved residues near to the N/C-terminal domains (blue box in **Figure 4**) are different between GH30_5 and GH30_11, although both subfamilies have the same conserved glutamic acid (E) residues involved in the nucleophile and acid/base catalytic mechanism. In addition, clearly different insertions were observed near some conserved residues in the sequences of GH30_5 and GH30_11 (green box in **Figure 4**), which may affect the catalytic pocket structure of enzymes and further lead to their different properties.

The strategies used in the GH30 study (**Chapter 3**), combining phylogenetic analysis with functional characterization, allowed us to discover not only novel enzymes but also novel subfamilies. The distinction of different subfamilies helps in the future to select candidates from CH30 with specific and defined functionalities. So far, out of 173 GH families in the CAZy database, only 5 families (i.e., GH5, GH13, GH16, GH30, and GH43) have been further divided into different subfamilies. Some families, such as GH1, GH2, GH3, GH31, GH32, and GH92, also have quite diverse activities, however, these activities are not further arranged in different subfamilies. New (sub)families are usually created based on amino acid sequence similarities. Inspired by the GH30 study, future work may be to create different subfamilies for other large CAZy families containing diverse activities by detailed phylogenetic analysis.

In **Chapter 4** and **5**, we performed a comparative characterization of multiple *P. subrubescens* enzymes from different families (GH10, GH11, GH51, GH54, and GH62). The results showed that the expansion of multiple enzymes within one family in *P. subrubescens* was accompanied by functional diversification of these enzymes. In addition, a previous study demonstrated that the expansion of α -galactosidases in GH27 and GH36 is accompanied by functional diversification of the enzymes [35]. Next, genome comparison of *P. subrubescens* with common industrial fungi showed that CAZyme expansion in *P. subrubescens* also occurred for several other CAZy families than those mentioned above, such as GH1, GH5, GH12, GH29, GH32, GH78, and GH93 (**Figure 5A**) [15, 36]. Such an expansion within one GH family based on one fungal genome, might hint at a diversification of functions within this family. The activities in these expanded CAZy families are mainly related to the degradation of hemicellulose, pectin and inulin. The fact that the expansion of *P. subrubescens* does not occur across all CAZy families may be driven by the native habitat of *P. subrubescens*, where *P. subrubescens* secretes an extensive but specific set of carbohydrate-active enzymes to degrade polysaccharides that appear in its habitat. However, whether functional diversification of enzymes is the norm in all expanded CAZy families of *P. subrubescens* is not clear. Studying additional CAZy families containing the expanded CAZyme-encoding genes in *P. subrubescens* will shed more light on this. Similarly, it is not clear whether this is a common feature among fungal species that have expansions in specific CAZy families.

The diverse CAZyme sets make *P. subrubescens* a highly interesting candidate for biotechnological applications. So far, only a handful of species (e.g., *A. niger*, *T. reesei*) have been used in industry, as mentioned at the beginning of this chapter. But besides general aspects, such as high protein production levels, low or no toxin production and fast growth behavior, another reason that makes fungi attractive for industrial applications is their ability to produce mixtures of plant biomass degrading enzymes. A previous study has demonstrated that *P. subrubescens* shows comparable potential to *A. niger* for enzymatic plant biomass saccharification [37]. Generally, the number of CAZyme-encoding genes in fungal genomes reflects, to some extent, their ability to degrade plant biomass [38-40]. Here I compared the completely sequenced genomes of ascomycete fungi in the MycoCosm database. Fungi containing a comparable number of CAZyme genes as *P. subrubescens* were listed, and these were from the Aspergilli genus (**Figure 5B**). Similar to *P. subrubescens*, these selected fungi harbored genes encoding a complex set of CAZymes that are probably involved in the degradation of major polysaccharides (cellulose, hemicellulose, and pectin) (**Figure 6**). In addition, multiple gene copies were observed in specific CAZy families in these selected fungi, such as in the families of GH5, GH11, GH12, GH27, GH36, GH51, GH78, and GH93, although the number of gene encoding CAZymes in these CAZy families was slightly lower than that of *P. subrubescens*.

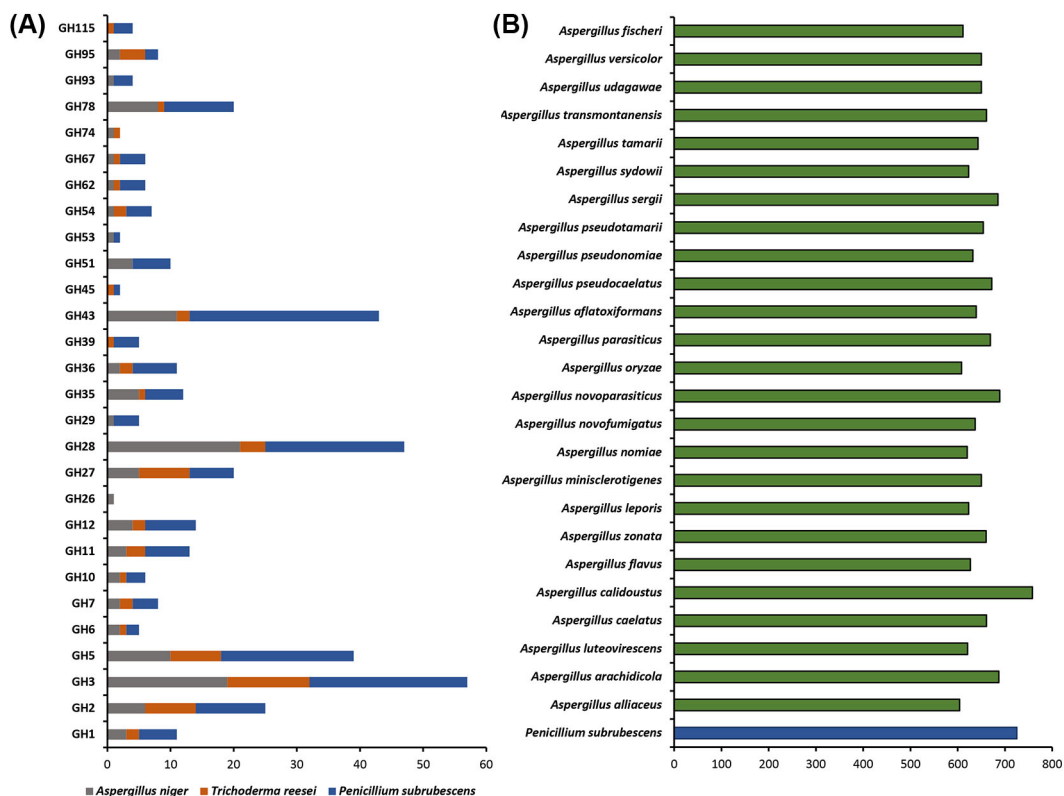


Figure 5 The number of CAZyme-encoding genes in ascomycete fungal genomes. (A) The number of genes in selected GH families related to plant biomass degradation in the genomes of common industrial fungi (data based on [15]). (B) CAZyme content of selected completely sequenced genomes of ascomycete fungi (data based on [22] and [36]).

(Figure 6). These selected fungi are expected to have excellent plant biomass degrading ability as they contain a rich set of genes encoding different CAZymes, and could be promising candidates for industrial applications. However, these predictions of their ability remain to be determined experimentally.

In contrast to the expansion of CAZyme-encoding genes in the fungi mentioned above, deletions of CAZyme-encoding genes are also observed in fungi. There are some notable examples, of which the ascomycete *T. reesei* is the most noteworthy one. *T. reesei* is a commonly used industrial cellulase producer but contains a relatively small number of cellulase-encoding genes in its genome [15, 41]. Additionally, *Rhizopus oryzae* is worth mentioning, because this fungus known for its high capability of plant cell wall degradation, but contains only a small number of CAZyme encoding genes in its genome [42]. Their strategy for plant biomass degradation seems to be focused on high production or high activity levels of a limited set of enzymes rather than on expanding their enzyme repertoire [43]. To discuss further, the expansion or deletion of CAZyme-encoding genes in fungal genomes may reflect their adaptive evolution in nature (ecological adaptation). Adaptive evolution due to the gain or loss of genes has been described previously in pathogens of plants [44, 45]. In plants, the immune system triggers defense responses after detection of specific pathogen proteins [46]. For pathogens, on the one hand, the deletions of genes encoding such detected proteins enhance adaptation to hosts to some extent. On the other hand, the expansion of gene encoding specific protein families can promote virulence on hosts. In brief, pathogens selectively secret proteins by expanding or deleting certain genes (families) in the genome to manipulate host (plant)

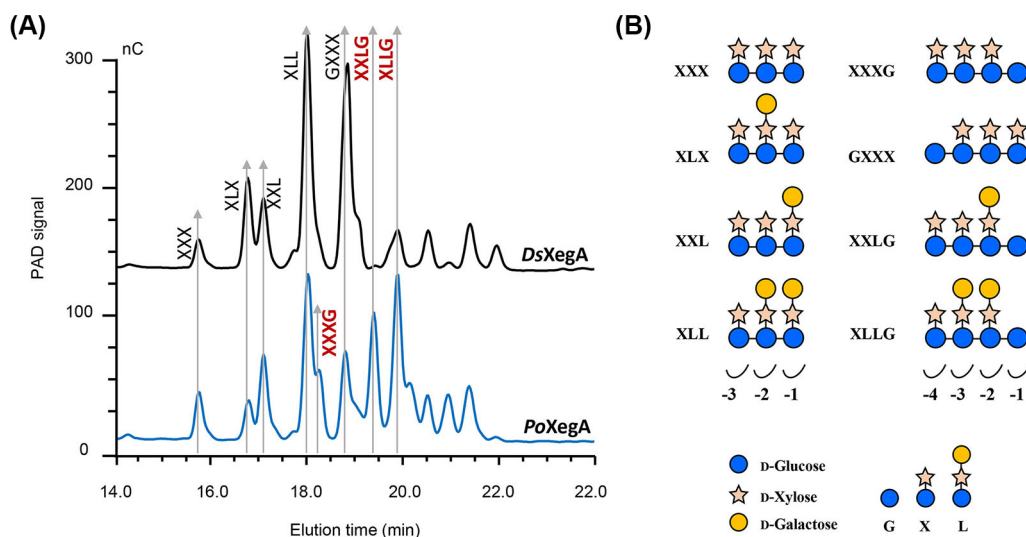


Figure 7 HPAEC elution patterns (A) of tamarind xyloglucan (TXG) oligosaccharides released from TXG by purified GH44 *DsXegA* and *PoXegA*, and schematic representation of the major released TXG oligosaccharides (B). Data is based on [48] (Chapter 6).

To be specific, as shown in **Figure 7**, a considerable amount of released xyloglucan oligosaccharides with an unsubstituted reducing terminal D-glucosyl residue (e.g., “XXXG”, “XXLG” and “XLLG”) can be observed in the GH44 *PoXegA*-TXG digest. Although these oligomers were also present in the GH44 *DsXegA*-TXG digest, xyloglucan oligosaccharides decorated with substitution at the reducing end, e.g., “XXL” and “GXXX” were more pronounced. Only a small amount of “XLLG” could be detected in the *DsXegA*-TXG digest. These results indicate that *DsXegA* may prefer to cleave glycosidic bonds next to branched glucopyranosyl residues in the main chain of TXG compared to *PoXegA*. This further reflects the difference in the key amino acids of their catalytic domain, especially in the -1 subsite. Additional TXG oligosaccharides with long main-chain and different side-chains need to be used as substrates to compare the difference of their hydrolysis efficiency. Furthermore, the effort to homology modeling analysis needs to be done in the future, which will help to understand catalytic mechanisms of GH44 fungal XEGs.

Compared to the other families (e.g., GH10, GH11) in this thesis, GH44 is a typically low-characterized plant biomass degrading family. The GH44 XEG activity we found in fungi has been previously reported in several bacterial species, suggesting a possible horizontal transfer of GH44 XEGs between fungi and bacteria [49-53]. In addition to GH44, other poorly characterized families related to plant biomass degradation, such as GH115, GH105, GH95, GH29, GH16, and GH7, which contain less than 5% of characterized fungal enzymes of the total number of enzymes, are also of interest [15]. Additional characterizations in low-characterized families, especially in terms of substrate specificity and biochemical properties, will not only help to improve the understanding of these families and thus to provide accurate functional annotation of genes from the corresponding families in fungal genomes, but also to develop better enzyme cocktails for biomass hydrolysis [15, 54].

Enzyme synergy for plant biomass degradation

A variety of enzymes with different activities have been discovered in this thesis, most of which can directly use polysaccharides as substrates and show high potential for the degradation of complex plant biomass. Previous studies have confirmed that synergy of different enzymes is essential

for the efficient conversion of plant biomass, which is the main strategy to improve the overall conversion of plant biomass in biotechnological applications. The actual synergistic degradation of total plant biomass was considered out-of-scope of this thesis. Nevertheless, based on the activities of enzymes found in this thesis (**Table 1**), I propose some ideas to study their synergism in future. Synergism between plant biomass-degrading enzymes is recognized as two different modes, namely homeosynergy and heterosynergy [55]. For polysaccharides that contains substituents on their backbone, it is expected that the enzymes degrading them exhibit both types of synergisms during their hydrolysis. Based on the final aims, I will mainly discuss the synergistic degradation of substituted polysaccharides.

Synergism to steer towards dimeric products for best prebiotic effect

In case different backbone-degrading enzymes are used for the production of different types of (substituted) oligosaccharides, their synergy might result in for example a more tailored range of DP (degree of polymerization), a larger variety of substitutions and/or in a lower enzyme dose needed. For example, for xylooligosaccharides (XOS), including xylobiose, production using synergistic combinations of (1) XLNs from different GH families and (2) XLN with XBH, the DP range obtained might be narrowed down in case of option 2 compared to 1. As previously mentioned, XLNs from different GH families exhibit different substrate specificities [56-58]. In **Chapter 5**, HPAEC-PAD results showed that *P. subrubescens* XLNs, even from the same family, showed diverse hydrolysis patterns. Therefore, combining XLNs with different hydrolysis behaviors usually leads to diversity and high production of shorter XOS. Among the XOS, the dimer xylobiose has been reported to have the highest prebiotic activity in *Bifidobacterium proliferation* and is most suitable for application in food industries [59-61]. The synergy between XLN and XBH may facilitate xylobiose production (**enzymes from Chapter 3, 5**).

Synergism to remove different side-chains more completely for higher recovery of substituent products and/or further subsequent backbone degradation

Different side chain-degrading enzymes are designed in part for the better recovery of substituent products, their cooperation would lead to a known and relatively simple product composition, and/or to a low separation cost. A previous study De Vries, Kester [62] provided insight into the synergistic effect of different side-chain degrading enzymes. Depending on the desired product, accessory enzymes are selectively added to the reaction. Synergistic combinations of (1) FAE and ABF or arabinoxylan arabinofuranohydrolase (AXH), and (2) α -glucuronidase (AGU) and ABF or AXH have been investigated [62]. These two pairs resulted in the higher production of ferulic acid and glucuronic acid, respectively, from insoluble pentosan, where the synergistic effect of the first combination was also manifested from the utilization of sugar beet pectin. This knowledge provides us with insights for setting up future synergy experiments using side-chain degrading enzymes from this thesis (**Chapter 2, 3, 4, Table 1**). For example, in **Chapter 2**, we characterized several enzymes in CE1 family, and it is worth mentioning the dual-active enzyme (FaeD) in CE1_2.1 (**Table 1**). This enzyme has been demonstrated to release acetic and ferulic acid from natural substrates (e.g., WAX). Future evaluation of the synergy of this dual enzyme with other accessory enzymes (e.g., ABF, AXH, AGU) towards glucuronoxylan (GX, *O*-acetyl-4-*O*-methylglucuroxylan), and glucuronoarabinoxylan (GAX, *O*-acetyl-arabino-4-*O*-methylglucuronoxylan) would be highly relevant (**Chapter 2, 4, Table 1**). It is worth mentioning that the synergy between different side chain-degrading enzymes can lower the substitution percentage of polysaccharides, which probably improve the hydrolysis efficiency of backbone-degrading enzymes (e.g., XLNs) [63-65]. This resulted synergy is named heterosynergy, which is different from homeosynergy mentioned above, i.e., synergy from different backbone-cleaving enzymes or from different side-chain cleaving enzymes.

Synergism to facilitate side-chain removal of substituted oligosaccharides for higher recovery of (substituent) products

To complement the mentioned homeosynergy (i.e., side-chain-cleaving enzyme remove substituents on polysaccharides that would hinder the liberation of oligosaccharides by backbone-cleaving enzymes), synergism also occurs when the main-chain-cleaving enzyme liberates substituted-oligosaccharides from polysaccharides for side-chain-cleaving enzymes [66]. This type of synergism might lead to a better efficiency of side-chain degrading enzyme (or a lower side-chain enzyme dose required), and a higher release of side-chain products. In **Chapter 2**, the synergistic action of FAE and XLN during the degradation of xylan substrates (i.e., WAX, wheat bran (WB) and corn xylooligosaccharides (CX)) was observed. Our results in **Chapter 2** showed that more ferulic acid was released from commercial XLN-treated xylan substrates than from untreated ones, which is consistent with previous studies that FAEs seems to prefer short substituted XOS generated by XLNs as substrates for cleavage of ester bonds to release ferulic acid. Using different *P. subrubescens* XLNs in **Chapter 5** instead of a commercial XLN may give unexpected results.

Last but not least, the study of the synergistic mechanisms between different enzymes in this thesis (**Table 1**) and the further incorporation of multiple enzyme activities in the reaction depending on the structure of the substrate is important to explore more efficient enzyme cocktails, which is one of the future implications of this thesis.

References

- Hollmann, F. and R. Fernandez-Lafuente, *Grand Challenges in Biocatalysis. Frontiers in Catalysis*, 2021. 1: p. 1.
- Narisetty, V., et al., *Valorisation of xylose to renewable fuels and chemicals, an essential step in augmenting the commercial viability of lignocellulosic biorefineries. Sustainable Energy & Fuels*, 2022. 6(1): p. 29-65.
- Zhang, X., K. Wilson, and A.F. Lee, *Heterogeneously catalyzed hydrothermal processing of C5–C6 sugars. Chemical Reviews*, 2016. 116(19): p. 12328-12368.
- Kango, N., U.K. Jana, and R. Choukade, *Fungal enzymes: sources and biotechnological applications, in Advancing frontiers in mycology & mycotechnology. Springer*, 2019: p. 515-538.
- El-Gendi, H., et al., *A Comprehensive Insight into Fungal Enzymes: Structure, Classification, and Their Role in Mankind's Challenges. Journal of Fungi*, 2022. 8(1): p. 23.
- Green, B.J. and D.H. Beezhold, *Industrial fungal enzymes: an occupational allergen perspective. Journal of Allergy*, 2011. 2011.
- Alcalde, M., et al., *Environmental biocatalysis: from remediation with enzymes to novel green processes. TRENDS in Biotechnology*, 2006. 24(6): p. 281-287.
- Kun, R.S., et al., *Developments and opportunities in fungal strain engineering for the production of novel enzymes and enzyme cocktails for plant biomass degradation. Biotechnology Advances*, 2019. 37(6): p. 107361.
- Salazar-Cerezo, S., et al., *CRISPR/Cas9 technology enables the development of the filamentous ascomycete fungus Penicillium subrubescens as a new industrial enzyme producer. Enzyme and Microbial Technology*, 2020. 133: p. 109463.
- Sharma, K.K., *Fungal genome sequencing: basic biology to biotechnology. Critical Reviews in Biotechnology*, 2016. 36(4): p. 743-759.
- Aylward, J., et al., *A plant pathology perspective of fungal genome sequencing. IMA Fungus*, 2017. 8(1): p. 1-15.
- Ahmad, M., et al., *Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. Applied Microbiology and Biotechnology*, 2014. 98(12): p. 5301-5317.
- Dilokpimol, A., et al., *Fungal glucuronoyl esterases: genome mining based enzyme discovery and biochemical characterization. New Biotechnology*, 2018. 40: p. 282-287.
- Dilokpimol, A., et al., *Fungal feruloyl esterases: functional validation of genome mining based enzyme discovery including uncharacterized subfamilies. New Biotechnology*, 2018. 41: p. 9-14.

15. de Vries, R.P. and M.R. Mäkelä, *Genomic and postgenomic diversity of fungal plant biomass degradation approaches*. **Trends in Microbiology**, 2020. 28(6): p. 487-499.
16. Cissé, O.H. and J.E. Stajich, *FGMP: assessing fungal genome completeness*. **BMC Bioinformatics**, 2019. 20(1): p. 1-9.
17. Sohn, J.I. and J.W. Nam, *The present and future of de novo whole-genome assembly*. **Briefings in Bioinformatics**, 2018. 19(1): p. 23-40.
18. Waterhouse, R.M., et al., *BUSCO applications from quality assessments to gene prediction and phylogenomics*. **Molecular Biology and Evolution**, 2018. 35(3): p. 543-548.
19. Zdobnov, E.M., et al., *OrthoDB v9. 1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs*. **Nucleic Acids Research**, 2017. 45(D1): p. D744-D749.
20. Hättenschwiler, S., A.V. Tiunov, and S. Scheu, *Biodiversity and litter decomposition in terrestrial ecosystems*. **Annual Review of Ecology, Evolution, and Systematics**, 2005. 36: p. 191-218.
21. Kivlin, S.N. and K.K. Treseder, *Initial phylogenetic relatedness of saprotrophic fungal communities affects subsequent litter decomposition rates*. **Microbial Ecology**, 2015. 69(4): p. 748-757.
22. Grigoriev, I.V., et al., *MycoCosm portal: gearing up for 1000 fungal genomes*. **Nucleic Acids Research**, 2014. 42(D1): p. D699-D704.
23. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013*. **Nucleic Acids Research**, 2014. 42(D1): p. D490-D495.
24. Dilokpimol, A., et al., *Screening of novel fungal Carbohydrate Esterase family 1 enzymes identifies three novel dual feruloyl/acetyl xylan esterases*. **FEBS Letters**, 2022.
25. Marcia, M.d.O., *Feruloylation in grasses: current and future perspectives*. **Molecular Plant**, 2009. 2(5): p. 861-872.
26. Ralph, J., et al., *Lignins: natural polymers from oxidative coupling of 4-hydroxyphenyl-propanoids*. **Phytochemistry Reviews**, 2004. 3(1): p. 29-60.
27. Hartley, R.D. and C.W. Ford, *Phenolic constituents of plant cell walls and wall biodegradability*. **American Chemical Society**. 1989: p. 137-139.
28. Lam, T., K. Kadoya, and K. Iiyama, *Bonding of hydroxycinnamic acids to lignin: ferulic and p-coumaric acids are predominantly linked at the benzyl position of lignin, not the β -position, in grass cell walls*. **Phytochemistry**, 2001. 57(6): p. 987-992.
29. Kylli, P., et al., *Antioxidant potential of hydroxycinnamic acid glycoside esters*. **Journal of Agricultural and Food Chemistry**, 2008. 56(12): p. 4797-4805.
30. Ou, J. and Z. Sun, *Feruloylated oligosaccharides: structure, metabolism and function*. **Journal of Functional Foods**, 2014. 7: p. 90-100.
31. Boo, Y.C., *p-Coumaric acid as an active ingredient in cosmetics: A review focusing on its antimelanogenic effects*. **Antioxidants**, 2019. 8(8): p. 275.
32. Dilokpimol, A., et al., *Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications*. **Biotechnology for Biofuels**, 2016. 9(1): p. 1-18.
33. Li, X., et al., *Glycoside Hydrolase family 30 harbors fungal subfamilies with distinct polysaccharide specificities*. **New Biotechnology**, 2022. 67: p. 32-41.
34. Kotake, T., et al., *Molecular cloning and expression in Escherichia coli of a Trichoderma viride endo-beta-(1 6)-galactanase gene*. **Biochemical Journal**, 2004. 377(3): p. 749-755.
35. Coconi Linares, N., et al., *Recombinant production and characterization of six novel GH27 and GH36 alpha-galactosidases from Penicillium subrubescens and their synergism with a commercial mannanase during the hydrolysis of lignocellulosic biomass*. **Bioresource Technology**, 2020. 295: p. 122258.
36. Peng, M., et al., *The draft genome sequence of the ascomycete fungus Penicillium subrubescens reveals a highly enriched content of plant biomass related CAZymes compared to related fungi*. **Journal of Biotechnology**, 2017. 246: p. 1-3.
37. Mäkelä, M.R., et al., *Penicillium subrubescens is a promising alternative for Aspergillus niger in enzymatic plant biomass saccharification*. **New Biotechnology**, 2016. 33(6): p. 834-841.
38. De Wit, P.J., et al., *The genomes of the fungal plant pathogens Cladosporium fulvum and Dothistroma septosporum reveal adaptation to different hosts and lifestyles but also signatures of common ancestry*. **PLoS Genetics**, 2012. 8(11): p. e1003088.
39. Benoit, I., et al., *Degradation of different pectins by fungi: correlations and contrasts between the pectinolytic enzyme sets identified in genomes and the growth on pectins of different origin*. **BMC Genomics**, 2012. 13(1): p. 1-11.

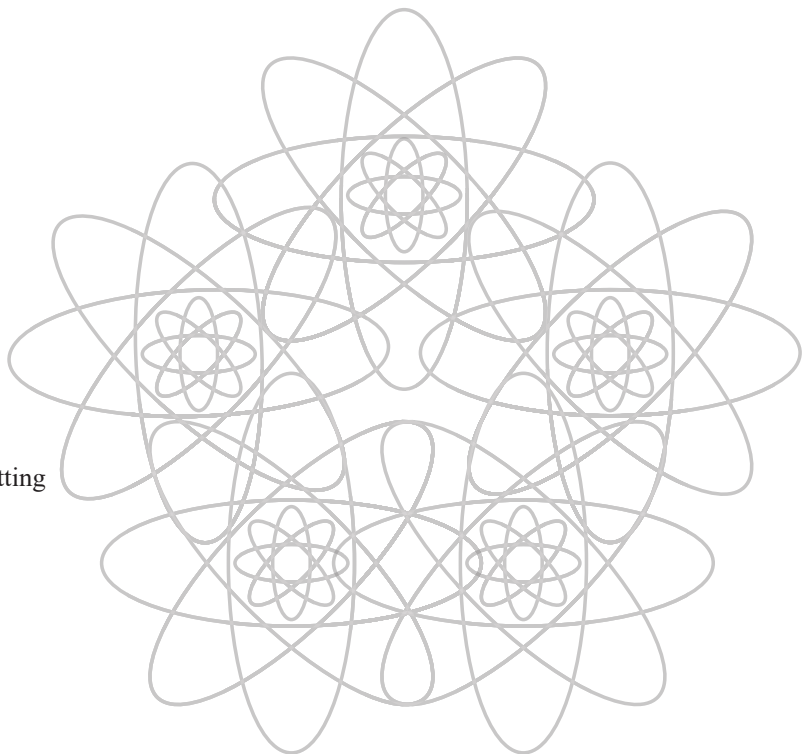
40. Suzuki, H., et al., *Comparative genomics of the white-rot fungi, Phanerochaete carnosa and P. chrysosporium, to elucidate the genetic basis of the distinct wood types they colonize.* **BMC Genomics**, 2012. 13(1): p. 1-17.
41. Martinez, D., et al., *Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).* **Nature Biotechnology**, 2008. 26(5): p. 553-560.
42. Battaglia, E., et al., *Carbohydrate-active enzymes from the zygomycete fungus Rhizopus oryzae: a highly specialized approach to carbohydrate degradation depicted at genome level.* **BMC Genomics**, 2011. 12(1): p. 1-12.
43. Bischof, R.H., J. Ramoni, and B. Seiboth, *Cellulases and beyond: the first 70 years of the enzyme producer Trichoderma reesei.* **Microbial Cell Factories**, 2016. 15(1): p. 1-13.
44. Hartmann, F.E. and D. Croll, *Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen.* **Molecular Biology and Evolution**, 2017. 34(11): p. 2808-2822.
45. Stukenbrock, E.H. and D. Croll, *The evolving fungal genome.* **Fungal Biology Reviews**, 2014. 28(1): p. 1-12.
46. Jones, J.D. and J.L. Dangl, *The plant immune system.* **Nature**, 2006. 444(7117): p. 323-329.
47. de Freitas Pereira, M., et al., *Secretome analysis from the ectomycorrhizal ascomycete Cenococcum geophilum.* **Frontiers in Microbiology**, 2018. 9: p. 141.
48. Sun, P., et al., *Fungal glycoside hydrolase family 44 xyloglucanases are restricted to the phylum Basidiomycota and show a distinct xyloglucan cleavage pattern.* **iScience**, 2022. 25(1): p. 103666.
49. Hirano, N., et al., *Cell-free protein synthesis and substrate specificity of full-length endoglucanase CelJ (Cel9D-Cel44A), the largest multi-enzyme subunit of the Clostridium thermocellum cellulosome.* **FEMS Microbiology Letters**, 2013. 344(1): p. 25-30.
50. Ariza, A., et al., *Structure and activity of Paenibacillus polymyxa xyloglucanase from glycoside hydrolase family 44.* **Journal of Biological Chemistry**, 2011. 286(39): p. 33890-33900.
51. Ravachol, J., et al., *Mechanisms involved in xyloglucan catabolism by the cellulosome-producing bacterium Ruminiclostridium cellulolyticum.* **Scientific Reports**, 2016. 6(1): p. 1-17.
52. Morañs, S., et al., *Enzymatic profiling of cellulosomal enzymes from the human gut bacterium, Ruminococcus champanellensis, reveals a fine-tuned system for cohesin-dockerin recognition.* **Environmental Microbiology**, 2016. 18(2): p. 542-556.
53. Warner, C.D., et al., *Kinetic characterization of a glycoside hydrolase family 44 xyloglucanase/endoglucanase from Ruminococcus flavefaciens FD-1.* **Enzyme and Microbial Technology**, 2011. 48(1): p. 27-32.
54. Drula, E., et al., *The carbohydrate-active enzyme database: functions and literature.* **Nucleic Acids Research**, 2022. 50(D1): p. D571-D577.
55. Moreira, L., *An overview of mannan structure and mannan-degrading enzyme systems.* **Applied Microbiology and Biotechnology**, 2008. 79(2): p. 165-178.
56. Christakopoulos, P., et al., *Antimicrobial activity of acidic xylo-oligosaccharides produced by family 10 and 11 endoxylanases.* **International Journal of Biological Macromolecules**, 2003. 31(4-5): p. 171-175.
57. Beaugrand, J., et al., *Impact and efficiency of GH10 and GH11 thermostable endoxylanases on wheat bran and alkali-extractable arabinoxylans.* **Carbohydrate Research**, 2004. 339(15): p. 2529-2540.
58. Broecker, J., et al., *The hemicellulose-degrading enzyme system of the thermophilic bacterium Clostridium stercorarium: comparative characterization and addition of new hemicellulolytic glycoside hydrolases.* **Biotechnology for Biofuels**, 2018. 11(1): p. 1-18.
59. Okazaki, M., S. Fujikawa, and N. Matsumoto, *Effects of xylooligosaccharide on growth of bifidobacteria.* **Bifidobacteria and Microflora**, 1990. 43(6): p. 395-401.
60. Moura, P., et al., *In vitro fermentation of xylo-oligosaccharides from corn cobs autohydrolysis by Bifidobacterium and Lactobacillus strains.* **LWT-Food Science and Technology**, 2007. 40(6): p. 963-972.
61. Choi, J.-H., et al., *Highly efficient recovery of xylobiose from xylooligosaccharides using a simulated moving bed method.* **Journal of Chromatography A**, 2016. 1465: p. 143-154.
62. De Vries, R.P., et al., *Synergy between enzymes from Aspergillus involved in the degradation of plant cell wall polysaccharides.* **Carbohydrate Research**, 2000. 327(4): p. 401-410.
63. Malgas, S., et al., *A mini review of xylanolytic enzymes with regards to their synergistic interactions during hetero-xylan degradation.* **World Journal of Microbiology and Biotechnology**, 2019. 35(12): p. 1-13.
64. Yang, C.-H. and W.-H. Liu, *Purification and properties of an acetylxyylan esterase from Thermobifida fusca.* **Enzyme and Microbial Technology**, 2008. 42(2): p. 181-186.
65. Zheng, F., et al., *A novel neutral xylanase with high SDS resistance from Volvariella volvacea: charac-*

- terization and its synergistic hydrolysis of wheat bran with acetyl xylan esterase. **Journal of Industrial Microbiology and Biotechnology**, 2013. 40(10): p. 1083-1093.*
66. Rosa, L., et al., *Characterization of a recombinant α -glucuronidase from *Aspergillus fumigatus*. **Fungal Biology**, 2013. 117(5): p. 380-387.*

Special

Appendix

English summary
Nederlandse samenvatting
Curriculum Vitae
List of publications
Acknowledgements



Unraveling the diversity within CAZy families related to hemicellulose degradation

Hemicellulose is the most abundant polysaccharide in nature after cellulose, according for 20-40% of agricultural waste streams, e.g., corn fiber, wheat straw, rice straw, and sugarcane bagasse. Its utilization is crucial for recycle and reuse of agricultural waste, with great environmental benefits, while also supporting the circular bio-economy that is currently being advocated. When it comes to the utilization of hemicellulose, much of our current attention is shifting to supplementing specific enzyme cocktails for effective bioconversion of hemicellulose into market-demanded products, rather than pursuing complete hydrolysis of hemicellulose with an army of enzymes. Custom-designed enzyme cocktails are receiving widespread attention, mainly because they can greatly reduce the costs of enzymatic hemicellulose deconstruction and subsequent product purification.

Fungi are highly efficient degraders of plant biomass. Most of them possess a rich set of genes in their genomes encoding hemicellulose-degrading enzymes that are likely to cooperate in hemicellulose deconstruction. Analysis of the differences of these enzyme sets in fungal genomes could provide clues for making efficient hemicellulose-degrading enzyme cocktails. However, most of genes in fungal genomes lack functional annotation. Predictions of such Carbohydrate Active enZymes (CAZymes) encoded by unknown genes based on the function of characterized fungal enzymes in CAZy database are not always reliable. This is largely caused by a low percentage (<15%) of characterized fungal enzymes compared to the total enzyme number in many CAZy families, e.g., CE1, GH35, GH43, GH115. This implies that the functional diversity within these CAZy families may be much higher than currently assumed. Therefore, experimental characterization of a wider set of CAZymes is crucial, not only to help refine fungal genome annotation and understand the hemicellulose-degrading ability of fungus, but also to reveal the diversity of CAZy families. An in-depth understanding of the diversity within a CAZy family can aid in the selection of optimal enzymes for different purposes and further upgrade existing enzyme mixtures.

Chapter 1 of this thesis describes the background and aim of my PhD project. For general background, I focused on xylan, the main component of hemicellulose. I first summarized the types, family classification, and substrate specificity of enzymes used for xylan degradation. Subsequently, the major approaches used to discover new xylanolytic enzymes were discussed, in particular the use of genome mining method was introduced. Furthermore, considering that the composition of hemicellulose is not limited to xylan, the applicability of genome mining for the discovery of other hemicellulases was briefly mentioned. Lastly, I overviewed the industrial applications of xylanolytic enzymes.

In **Chapter 2** and **3**, we mainly characterized enzymes from poorly characterized CAZy families to deepen our understanding of them. In **Chapter 2**, we selected six ascomycete candidates from poorly studies clades of CE1_SF1 and CE1_SF2 in the CE1 phylogenetic tree, and successfully characterized four of them. We found that a CE1_SF2 enzyme was active towards all tested model substrates (e.g., MUB-acetate, methyl *p*-coumarate, methyl caffeate, methyl ferulate, methyl sinapate, ethyl *p*-coumarate, ethyl ferulate and chlorogenic acid), while three CE1_SF1 enzymes showed activity only towards MUB-substrates. Additionally, an enzyme from CE1_SF2 was able to release acetic acid and ferulic acid from plant biomass (e.g. wheat arabinoxylan (WAX), wheat bran, sugar beet pectin), whereas enzymes from CE1_SF1 could only release acetic acid from WAX or/and corn oligosaccharides. These results suggests that CE1_SF1 contains acetyl xylan esterases (AXEs), while CE1_SF2 also contains feruloyl esterases (FAEs). The enzyme with dual activities (FAE/AXE) from CE1_SF2 partially reflects the evolutionary relationship between CE1_SF1 and CE1_SF2.

In **Chapter 3**, we evaluated the diversity within GH30 family by characterization of eleven enzymes from different fungal subfamilies. Our results confirm these subfamilies possess enzymes with distinct substrate specificities. In addition to the reported endohydrolase activities of each subfamily, e.g., β -(1 \rightarrow 6)-endoglucanases of GH30_SF3, β -(1 \rightarrow 6)-endogalactanases of GH30_SF5, and endoxylanases of GH30_SF7, fungal GH30 subfamilies contain novel disaccharides hydrolase

activities, e.g., β -(1 \rightarrow 6)-galactobiohydrolases in GH30_SF5 and xylobiohydrolases in GH30_SF7. Furthermore, our study led to the discovery of GH30_11, a new fungal GH30 subfamily containing β -(1 \rightarrow 6)-galactobiohydrolases. Taken together, our findings suggests that fungal GH30 enzymes have huge potential for releasing ‘short’ non-digestible di- and oligosaccharides.

In **Chapter 4** and **5**, we aimed to understand the added value of CAZyme expansions in *Penicillium subrubescens* by characterization of enzymes from expanded CAZy families related to the degradation of hemicellulose. In this thesis, we comparatively analyzed nine α -L-arabinofuranosidases (ABFs) from GH51, GH54, and GH62 in **Chapter 4**. We found that these ABFs had similar physical characteristics, but exhibited diverse properties in terms of pH, temperature, substrate specificity, and activity levels, especially their activities on natural substrates (wheat arabinoxylan, sugar beet pectin, arabinan and debranched arabinan). In **Chapter 5**, we describe ten endoxylanases (XLNs) from GH10 and GH11 using beechwood xylan and WAX as substrates. Clear differences in specific activity and hydrolysis pattern were observed among different *P. subrubescens* XLNs. Based on the homology modelling, we inferred that the diverse product patterns of WAX hydrolysis by *P. subrubescens* XLNs results from the different features in their catalytic domain. Taken together, our results in **Chapter 4** and **5** suggest that the expansion of ABFs and XLNs in the *P. subrubescens* genome is accompanied by functional diversity. These findings likely reflect an evolutionary adaptation of *P. subrubescens* that provides a diverse and flexible enzymatic toolbox for synergistic degradation of hemicellulose in its natural habitat. This makes *P. subrubescens* and others with similar expansions highly interesting candidates for biotechnological applications.

In **Chapter 6**, we characterized two fungal endoglucanases (XEGs) in GH44, a family that previously contained only bacterial XEGs, and compared their xyloglucan cleavage patterns with fungal XEGs from other CAZy families. Our results indicated that these two fungal GH44 XEGs showed highly similar biochemical properties, however, they had distinct xyloglucan cleavage patterns compared to fungal XEGs from GH12 and GH74. They were not hindered by substitution of neighboring D-glucosyl units and generated unique xyloglucan oligosaccharide products (e.g., “XXXG-type”, “XXX-type”, “GXXXG-type”).

Taken together, the results described in this thesis not only demonstrate the feasibility of the genome mining strategy for the discovery of novel enzymes, but also unravel the functional diversity within and between CAZy families. In the final **Chapter 7**, I discuss considerations for using genome mining strategy by researchers working in related areas. In addition, I provide some ideas for possible future extended research aspects based on results in this thesis. Lastly, to update or make enzyme cocktails that efficiently degrade plant biomass, I present some ideas on how to use the enzymes found in this study to set up synergistic experiment.



Het ontrafelen van de diversiteit binnen CAZy families met betrekking tot de afbraak van hemicellulose

Hemicellulose is het meest voorkomende polysaccharide in de natuur na cellulose, en omvat 20-40% van de landbouwafvalstromen, bijvoorbeeld maïsvezel, tarwestro, rijststro en suikerrietbagasse. Het gebruik ervan is cruciaal voor recycling en hergebruik van landbouwafval, met grote voordelen voor het milieu, terwijl het ook de circulaire bio-economie ondersteunt die momenteel wordt bepleit. Als het gaat om het gebruik van hemicellulose, verschuift veel van onze huidige aandacht naar het aanvullen van specifieke enzymcocktails voor effectieve bioconversie van hemicellulose in producten die op de markt worden gevraagd, in plaats van het nastreven van volledige hydrolyse van hemicellulose met een brede set enzymen. Op maat ontworpen enzymcocktails krijgen veel aandacht, vooral omdat ze de kosten van enzymatische deconstructie van hemicellulose en daaropvolgende opzuivering van de producten aanzienlijk kunnen verlagen.

Schimmels zijn zeer efficiënte afbrekers van plantaardige biomassa. De meeste van hen bezitten een grote set genen in hun genomen die coderen voor hemicellulose afbrekende enzymen die waarschijnlijk samenwerken bij de deconstructie van hemicellulose. Het analyseren van verschillend in deze sets van enzymen in schimmelgenomen zou aanwijzingen kunnen geven voor het maken van efficiënte hemicellulose afbrekende enzymcocktails. De meeste genen in het genoom van schimmels missen echter functionele annotatie. Voorspellingen van dergelijke koolhydraat-actieve enzymen (CAZymes) die worden gecodeerd door onbekende genen op basis van de functie van gekarakteriseerde schimmelenzymen in de CAZy database zijn niet altijd betrouwbaar. Dit wordt grotendeels veroorzaakt door een laag percentage (<15%) gekarakteriseerde schimmelenzymen in vergelijking met het totale aantal enzymen in veel CAZy families, zoals CE1, GH35, GH43, GH115. Dit impliceert dat de functionele diversiteit binnen deze CAZy families veel hoger kan zijn dan momenteel wordt aangenomen. Daarom is experimentele karakterisering van een bredere reeks CAZymes cruciaal, niet alleen om de annotatie van schimmelgenomen te verbeteren en het hemicellulose afbrekende vermogen van schimmels te begrijpen, maar ook om de diversiteit van CAZy families te onthullen. Een diepgaand begrip van de diversiteit binnen een CAZy familie kan helpen bij de selectie van optimale enzymen voor verschillende toepassingen en het verder verbeteren van bestaande enzymmengsels.

Hoofdstuk 1 van dit proefschrift beschrijft de achtergrond en het doel van mijn promotieonderzoek. Voor de algemene achtergrond heb ik me vooral gericht op xylan, het hoofdbestanddeel van hemicellulose. Ik heb eerst de typen, familieclassificatie, en substraatspecificiteit van enzymen samengevat die worden gebruikt voor xylan afbraak. Vervolgens werden de belangrijkste benaderingen besproken die zijn gebruikt om nieuwe xylanolytische enzymen te ontdekken, met name het gebruik van de genoom mining methode. Aangezien de samenstelling van hemicellulose niet beperkt is tot xylan, werd ook de toepasbaarheid van genoom mining voor de ontdekking van andere hemicellulases kort genoemd. Ten slotte heb ik een overzicht gegeven van de industriële toepassingen van xylanolytische enzymen.

In **Hoofdstuk 2** en **3** hebben we voornamelijk enzymen gekarakteriseerd uit slecht gekarakteriseerde CAZy families om ons begrip ervan te verdiepen. In **Hoofdstuk 2** selecteerden we zes ascomycete kandidaten uit slecht bestudeerde takken van CE1_SF1 en CE1_SF2 in de CE1 fylogenetische boom, en we karakteriseerden vier daarvan. We ontdekten dat een CE1_SF2 enzym actief was op alle geteste modelsubstraten (bijv. MUB-acetaat, methyl-*p*-coumaraat, methylcaffeaat, methylferulaat, methylsinaapaat, ethyl-*p*-coumaraat, ethylferulaat en chlorogeenzuur), terwijl drie CE1_SF1 enzymen alleen activiteit vertoonden tegen MUB-substraten. Bovendien was een enzym van CE1_SF2 in staat om zowel acetaat als ferulazuur vrij te maken uit plantaardige biomassa (bijvoorbeeld tarwe arabinoxylan (WAX), tarwe zemelen, suikerbiet pectine), terwijl enzymen uit CE1_SF1 alleen acetaat uit WAX en/of maïs oligosacchariden konden vrijmaken. Deze resultaten suggereren dat CE1_SF1 acetyl xylan esterasen (AXEs) bevat, terwijl CE1_SF2 ook feruloyl esterasen (FAEs) bevat. Het enzym met dubbele activiteit (FAE/AXE) van CE1_SF2 weerspiegelt gedeeltelijk de evolutionaire relatie tussen CE1_SF1 en CE1_SF2.

In **Hoofdstuk 3** hebben we de diversiteit binnen de GH30 familie geëvalueerd door elf enzymen uit verschillende subfamilies van schimmels te karakteriseren. Onze resultaten bevestigen dat deze subfamilies enzymen bezitten met verschillende substraat specificiteiten. Naast de gerapporteerde endohydrolase activiteiten van elke subfamilie, bijv. β -(1 \rightarrow 6)-endoglucanase van GH30_SF3, β -(1 \rightarrow 6)-endogalactanase van GH30_SF5 en endoxylanase van GH30_SF7, bevatten schimmel GH30 subfamilies nieuwe disaccharide hydrolase activiteiten, bijvoorbeeld β -(1 \rightarrow 6)-galactobiohydrolase in GH30_SF5 en xylobiohydrolase in GH30_SF7. Bovendien leidde onze studie tot de ontdekking van GH30_11, een nieuwe schimmel GH30 subfamilie die β -(1 \rightarrow 6)-galactobiohydrolase bevat. Al met al suggereren onze bevindingen dat GH30 enzymen van schimmels een enorm potentieel hebben voor het vrijmaken van ‘korte’ niet-verteerbare di- en oligosacchariden.

In **Hoofdstuk 4** en **5** wilden we de toegevoegde waarde van CAZyme expansies in *Penicillium subrubescens* begrijpen door enzymen uit uitgebreide CAZy families te karakteriseren die verband houden met de afbraak van hemicellulose. In dit proefschrift hebben we negen α -L-arabinofuranosidasen (ABFs) van GH51, GH54 en GH62 vergeleken in **Hoofdstuk 4**. We ontdekten dat deze ABFs vergelijkbare fysieke kenmerken hadden, maar verschillende eigenschappen vertoonden in pH, temperatuur, substraat specificiteit, en activiteitsniveaus, vooral met betrekking tot hun activiteit op natuurlijke substraten (tarwe arabinoxylan, suikerbiet pectine, arabinan en vertakt arabinan). In **Hoofdstuk 5** beschrijven we tien endoxylanasen (XLNs) van GH10 en GH11 met beukenhout, xylan en WAX als substraten. Er werden duidelijke verschillen in specifieke activiteit en hydrolyse patroon waargenomen tussen verschillende *P. subrubescens* XLNs. Op basis van homologie modellering concludeerden we dat de diverse productprofielen van WAX hydrolyse door *P. subrubescens* XLNs het gevolg zijn van de verschillen in hun katalytische domein. Samengevat suggereren onze resultaten in **Hoofdstuk 4** en **5** dat de uitbreiding van ABFs en XLNs in het *P. subrubescens* genoom gepaard gaat met functionele diversiteit. Deze bevindingen weerspiegelen waarschijnlijk een evolutionaire aanpassing van *P. subrubescens* die een diverse en flexibele enzymatische gereedschap set biedt voor synergetische afbraak van hemicellulose in zijn natuurlijke habitat. Dit maakt *P. subrubescens* en anderen met vergelijkbare uitbreidingen zeer interessante kandidaten voor biotechnologische toepassingen.

In **Hoofdstuk 6** hebben we twee schimmel endoglucanasen (XEGs) in GH44 gekarakteriseerd, een familie die voorheen alleen bacteriële XEGs bevatte, en hun xyloglucan splitsingspatronen vergeleken met schimmel XEGs van andere CAZy families. Onze resultaten gaven aan dat deze twee schimmel GH44 XEGs zeer vergelijkbare biochemische eigenschappen vertoonden, maar hun xyloglucan splitsingspatronen verschilden van schimmel XEGs van GH12 en GH74. Ze werden niet gehinderd door substitutie van aangrenzende D-glucosyl eenheden en genereerden unieke xyloglucan-oligosaccharide producten (bijv. “XXXG-type”, “XXX-type”, “GXXXG-type”).

Alles bij elkaar genomen demonstreren de resultaten beschreven in dit proefschrift niet alleen de haalbaarheid van de genoom mining strategie voor de ontdekking van nieuwe enzymen, maar ontrafelen ze ook de functionele diversiteit binnen en tussen CAZy families. In het laatste **hoofdstuk 7** bespreek ik overwegingen voor het gebruik van genoom mining strategie door onderzoekers die in verwante gebieden werken. Daarnaast geef ik enkele ideeën voor mogelijke toekomstig verder onderzoek op basis van de resultaten in dit proefschrift. Ten slotte, om enzymcocktails te verbeteren of te maken die de biomassa van planten efficiënt afbreken, presenteer ik enkele ideeën over hoe de enzymen die in deze studie zijn gevonden kunnen worden gebruikt om een synergetisch experiment op te zetten.



Xinxin Li was born on June 29th, 1992, in Henan, China. After high school, she enrolled at Henan Institute of Science and Technology in Henan, China. As part of her Bachelor study, she did a short internship as a technician at CP GROUP (正大集团) in Nanyang, Henan. In 2015, she obtained her Bachelor-level diploma in Animal Science. She continued with Master study in 2015 in the Chinese Academy of Agricultural Sciences, with specialization in Animal Nutrition and Feed Sciences. She completed her master thesis entitled “Effects of glycosylation on the enzymatic properties of β -Glucosidase Bgl3A expressed in *Pichia pastoris*” at Institute of Feed Research, and received her Master degree in 2018. In October of the same year, she registered as a PhD student at Utrecht University, and conducted her PhD research in the Fungal Physiology group at the Westerdijk Fungal Biodiversity Institute (Utrecht, The Netherlands), under the supervision of Prof. Dr. Ronald P. de Vries and Dr. Mirjam Kabel. The results of her PhD research are described in this thesis on “Unraveling the diversity within CAZy families related to hemicellulose degradation”.

Li, Xinxin, Dimitrios Kouzounis, Mirjam A. Kabel, Ronald P. de Vries, and Adiphol Dilokpimol. “Glycoside Hydrolase family 30 harbors fungal subfamilies with distinct polysaccharide specificities.” *New Biotechnology* 67 (2022): 32-41.

Li, Xinxin, Adiphol Dilokpimol, Mirjam A. Kabel, and Ronald P. de Vries. “Fungal xylanolytic enzymes: Diversity and applications.” *Bioresource Technology* 344 (2022): 126290.

Li, Xinxin, Dimitrios Kouzounis, Mirjam A. Kabel, and Ronald P. de Vries. “GH10 and GH11 endoxylanases in *Penicillium subrubescens*: comparative characterization and synergy with GH51, GH54, GH62 α -L-arabinofuranosidases from the same fungus.” *New Biotechnology* (2022).

Sun, Peicheng*, **Xinxin Li***, Adiphol Dilokpimol, Bernard Henrissat, Ronald P. de Vries, Mirjam A. Kabel, and Miia R. Mäkelä. “Fungal glycoside hydrolase family 44 xyloglucanases are restricted to the phylum Basidiomycota and show a distinct xyloglucan cleavage pattern.” *iScience* 25, no. 1 (2022): 103666.

Li, Xinxin*, Kelli Griffin*, Sandra Langeveld, Matthias Frommhagen, Emilie N. Underlin, Mirjam A. Kabel, Ronald P. De Vries, and Adiphol Dilokpimol. “Functional validation of two fungal subfamilies in carbohydrate esterase family 1 by biochemical characterization of esterases from uncharacterized branches.” *Frontiers in Bioengineering and Biotechnology* (2020): 694.

Coconi Linares, Nancy, **Xinxin Li**, Adiphol Dilokpimol, and Ronald P. de Vries. “Comparative characterization of nine novel GH51, GH54 and GH62 α -L-arabinofuranosidases from *Penicillium subrubescens*.” *FEBS Letters* 596, no. 3 (2022): 360-368.

Dilokpimol, Adiphol, Bart Verkerk, **Xinxin Li**, Annie Bellemare, Mathieu Lavallee, Matthias Frommhagen, Emilie Nørmølle Underlin et al. “Screening of novel fungal Carbohydrate Esterase family 1 enzymes identifies three novel dual feruloyl/acetyl xylan esterases.” *FEBS Letters* (2022).

*These authors contributed equally to this work.



All good things must come to an end.

Geoffrey Chaucer

Although I am reluctant to say “goodbye”, I realized that my PhD journey is drawing to an end when I started writing the final chapter “Acknowledgement” for my thesis. Standing at the tail of graduation, looking back, this PhD journey has been full of ups and downs, sweetness and bitterness. But what I can say for sure is that I really enjoyed this journey, and it is destined to be the most important and sweetest memory of my study life! I’ve always considered myself lucky because I have received help, support and care from so many people in so many different ways over the past four years! I was fortunate enough to apply for funding from the China Scholarship Council in 2018. Without their financial support, I would never have had the opportunity to live and study abroad for four years. The greatest fortune is that I realize that I have always been surrounded by a group of warm and kind people (my supervisors, colleagues, friends, family). Without their tremendous help, care and encouragement, I would not have been able to complete my PhD and submit this thesis. I would like to sincerely THANK all of YOU for your support, contributions, for your company and all the good moments along this journey. I’m sure I will miss a few names, but I would also like to thank you, no doubts.

First and foremost, I would like to express my sincere thanks to my major supervisor and promoter, **Ronald de Vries**. Through our first skype meeting (18 January 2018), you knew that my English was poor (I was acting particularly silly and made the interview more or less awkward), but you were still willing to give me the opportunity to join Fungal Physiology and support my application for the China Scholarship Council (CSC) grant. You could not image how excited I was when I emailed you that I was granted by CSC, especially when I received your email telling me I could start my PhD study on 1 October 2018. Thank you for your willingness to take the risks and accept me as one of your PhD students. I really admire your enthusiasm for research, which is very motivating! During these four-year studies, I have been troubled by experimental results many times, but after each discussion with you, you could always find solutions to my problems, got me on the right track, and reduced my pressure. You are such an amazing supervisor! Your comforting phrases during every PhD update meeting, such as “We are getting there”, and “This could be a new paper”, always gave me new energy! Thank you for your guidance, support, and encouragement these years! I also appreciate your insightful feedback on each of my manuscripts and presentations, which inspire me and help me grow as a research scientist. I would also like to thank you for organizing different activities (dinners, hiking, cycling, BBQ, and upcoming wadlopen and horse riding), and sending “happy birthday” on my birthday, which all made me feel so warm! If there were no limits, I thought I could write a few pages to express my gratitude to you over the years. Because there are so many things you organized that are not mentioned, such as amazing dinner with other groups in Rome (ECFG), Christmas dinner in your new house (with Natalia), Project dinner in Utrecht city center (with Mirjam), and many farewell parties (offline and online). Thank you again for everything you have done, it has made my life in the Netherlands beautiful and fulfilling! I will keep improving my English and hope to work with you again in the near future. As you always said, your door was always open for us. Similar to Tania, consider mine open for you as well, your emails/meetings are welcome at any time, and also look forward to your coming to China☺!

Secondly, I would like to thank my first co-promotor, **Adiphol Dilokpimol**. Adiphol, a great daily supervisor! I cannot thank you enough, in particular during October 2018 - January 2011. You were my first support at the WesterdijkInstitute when I was new and could barely express myself in English. You pushed me to listen, to speak, and also you were a person that corrected my pronunciation from time to time. Thank you for your constant encouragement in my English learning (English learning is always on the way). I would also like to thank you for all your patience in teaching me how to design/conduct experiments, analyze results, write reports/outlines/manuscripts, and many other little things. You are detail-oriented, and you can find even the tiniest typo or punctuation in my manuscripts. You used to be a wonderful helper in lab, and help us in many ways, thank you for

always being there to help me every time I needed it! You could always come with up many different solutions to varied issues I met during my work, and most of them worked! Also thank you for the book (“The things you can see only when you slow down”) you gave me before you left. Patience is something I need to spend a lifetime learning. Thank you again for all the good times you brought! As Ronnie said, you are the best supervisor a PhD student can have! I totally agree with it! Good luck with your work in Finland. I do have a trip planned to Espoo, Finland after my defense, and hopefully we can have a lunch or dinner together☺.

Special thanks also to my second co-promotor from Wageningen University, **Mirjam Kabel**. You were one of the key people who helped me the most throughout my whole research (you have contributed to every chapter). It was a great pleasure to have your guidance during the last two years of my PhD studies. Whenever Ronald and I were unsure about some results/ideas, we concluded by saying “Let’s ask Mirjam” “Let’s wait for Mirjam’s reply”. And indeed, you always provided us with new perspectives to analyze and sort out our questions. Thank you for your critical reading and insightful feedback of the manuscripts, which certainly made manuscript more interesting and suitable for publications. I have learned a lot from your feedback and revisions, which will surely help me in the future. I really enjoyed the dinner we had in Utrecht city center with Ronald, still remembered you talked about XiAn ☺, Ronald talked about Shanghai ☺, of course, we all also talked about my PhD project. Hope one day you will come to China again and I will show you around Henan, Beijing and other places. I would also like to thank **Dimitris Kouzounis** and **Peicheng Sun (培成)** in your lab. Thank you a lot for helping to prepare/analyze a large number of HPAEC/HPSEC samples, and also for being available to discuss the results with us! I would also like to thank you for your comments and revisions to manuscripts. Without your valuable input, Chapter 3, Chapter 5, and Chapter 6 would not be published! Thank you for your great contributions. Dimitris, thank you also for your taking care of me and Adiphol in your lab ☺. I wish you all the best finishing up your PhD this year ☺. Peicheng, thank you also for answering my stupid questions about GH44 manuscript ☺. I hope everything will work out for your post-doc and hopefully we can meet in China one day.

I would also like to thank **Miia Mäkela** from Helsinki University. I appreciate all of your contributions in GH44 family! Without your great expertise, support, and detailed revision, GH44 manuscript of Chapter 6 would not be published. **Nancy Coconi Linares**, I did not have the opportunity to meet you, but thank you for your kind cooperation in ABF manuscript. Without your tremendous input, Chapter 4 would not be possible! **Kelli Griffin**, thank you for your great contributions in Chapter 2, it is immeasurable! I really enjoyed working with you. I have never met a girl as much energy and confidence as you until 2018! Similar thanks also to other co-authors (**Sandra Langeveld**, **Matthias Frommhagen**, **Emilie N. Underlin**, **Bernard Henrissat**), I really appreciate all your comments/advices/revisions on my manuscripts! Without all your contributions, manuscripts from Chapter 2, Chapter 4, and Chapter 6 would not be possible! I would also like to thank **Jos Houbraken**. Thank you for your help in checking the names of *Aspergillus* species in Chapter 7!

I am also grateful to the members of the defense committee, **Prof. dr. J.C.M. Sjef Smeekens** (Utrecht University), **Prof. dr. H.A. Schols** (Wageningen University), **Prof. dr. J. Delcour** (Leuven University), **Dr. ir. N.N.M.E. Van Peij** (DSM), **Prof. dr. A. Tsang** (Concordia University), **Prof. dr. G.J.P.H. Boons** (Utrecht University). Thank you for taking your time to attend my defense and, most importantly, for your reading and evaluating my thesis.

Now, my deepest thanks and gratitude to WesterdijkInstitute and FP group, where I stayed for four years! The order in which you are mentioned does not change my gratitude to all of you!

Sumitha, a nice post-doc in lab. We met in 2018, but at that time you were pregnant, and had a long holiday. I did not have the opportunity to learn from you. But after you came back, I indeed learned a lot from you. Thank you for always answering my questions, from experimental principles to protein structures. I really enjoyed discussing with you, though sometimes in loud and strange ways ☺. I missed our conversations about job/life in the lab, especially during our lab cleaning day. I hope we will have more conversations in the future, you are such a warm person! **Agata**, the ever-energetic person! I really enjoyed sharing the office with you, Roland and Ronnie in 2019. Once I looked back



at you or your computer, you were always willing to talk/share with me. Thank you for being such a kind office mate, and also for all good moments you gave. It was so nice to meet up with you again in 2021. I wish you all the best finishing up your PhD, and also the best of luck in your future job :) **Astrid**, hard-working lab steward! You have been helping us all the time! Thank you for being so supportive in the lab, and many thanks for your help in ordering many chemicals/enzymes/kits and reminding me when they arrived immediately! The best of luck with your PhD!

Natalia, it was a pleasure working with you in the lab. I still remembered my awkward experience in the biochemical lab, thank you for your company and chat! I would also like to thank you for your PowerPoint and for sharing your specially scientific icons! I enjoyed the dinner with you and Ronald in his new house. And I also appreciated the conversations we had via TEAMS when I was in a sad mood (you were back in Spain by that time). I wish you all the best in your PhD. Many thank you for your support, care and help. **Sonia**, I enjoyed going to English courses with you every Thursday (maybe Tuesday?) evening ☺. I felt pretty safe when I stayed with you in the class (I was afraid of the teacher asking me questions)! I also missed that very rushed dinner in my apartment, it was full of fun (even though we were still late to class)! **Ola**, I really enjoyed the time we spent sitting in Fishtank, haha ☺. I know you think that place is too noisy. Thank you for always being available to chat and share with me, and your encouragement and support are much appreciated! Good luck with your Master!

Roland and Ronnie, my former office mates. You guys got me into good work habits, STAY AWAY FROM PHONE! You are my examples to follow and how to become such a productive PhD! I really enjoyed working with you in the same office. Roland, good luck with your PhD defense! **Sandra GC** and **Tania**, thank you for your great help in lab, and it was a lot fun working with you. It is amazing how you always have something interesting to share in the lab or at lunchtime! Roland also! **Luca** and **Alessia**, you are such a hard-working students! I am so grateful for our collaboration on the alpha-glucuronidase project! I definitely have learned a lot from supervising you. I wish you all the best for your future!

Mao (彭茂师兄), Jiali (佳丽), Jiajia (佳佳), Li (许力), and Dujuan (杜娟), my Chinese colleagues in FP group ☺. Thank you for your big help in many different ways! You started as my colleagues and ended as my friends! 彭茂师兄, 谢谢您在科研方面给予的建议, 每次组会您的提问都对我特别有启发, 让我能不断完善我的“科研故事”! 也特别感谢师兄在回答我“转录学, 蛋白组学”问题时的耐心! 更感谢师兄对我的一些中文材料的细心修改! 希望师兄未来能够申请更多项目, 发表更多高质量的文章! 佳丽, 你真的是一个很酷的人, 很专注于自我, 这一点很值得我学习! 谢谢佳丽在日常实验以及准备答辩事宜前后给予的莫大帮助, 希望我们两个都能顺利通过答辩, 也希望你能顺利找到喜欢的工作! 佳佳, 写得一手好代码的生信博士生! 和你讨论科研的感觉好棒, 但是和你逛街的感觉更棒!)! 回国后我一定会特别怀念和你逛街, 聊天的日子:)。希望你科研顺利, 明年顺利毕业, 也祝福你和你的毛先生生长长久久! 许力, 刚开始对你了解不多, 但是自从我们今年“南法之行”之后, 在我心目中你已经是“许牛牛”了:)。期待与你接下来的周末“比利时游”(和佳佳, 圆圆师姐)和“德国游”(和佳丽, 圆圆师姐)! 你是一个可爱有趣的妹纸, 希望在荷兰的接下来的两年内, 你能变得外向一些, 慢慢变成“社牛”! 杜鹃, 谢谢你教我Crispr cas9技术! 你对生活积极乐观的态度真的很值得学习。希望你和许力科研顺利, 后年能够一起毕业! 感恩遇见, 希望有机会能与你们每一个人再相见!

I would also like to thank all my other colleagues in FP group during these years: **Ad, Paul, Bas, Kimberly, Mar, Zoey, Melva, Abi, Jochem, Chendo, Raquel, Alejandra**...It was a great memory to work with you. Thank you for the great atmosphere in and around the lab, for all the fun moments, and most of all for all the help that you gave. Thanks also to the people from other groups, **Martin, Roya, Olga, Caroline, Trung**, it was a pleasure to chat with you.

Pedro, the director of Westerdijk Fungal Biodiversity Institute, thank you for making the institute such a great place to work. I have never seen such a energetic director like you! I really admire your passion for working with fungi, and I absolutely miss your interesting speeches on various occasions (New year, Christmas, Monday seminars, etc)!

Manon, the management Assistant of Westerdijk Fungal Biodiversity Institute, you are definitely the sun of our institute, thank you for being so supportive and helpful whenever we needed! Without your Adobe InDesign course, the formatting of my thesis would have been a mess, and could not have been fully completed in a short time!

My Chinese corner in WesterdijkInstitute©! Thank you for being there as great mentors/friends during my PhD, and for all the fun and good times you created! **Dongmei Shi (冬梅姐)**, **Liang Zhang (亮姐)**, **Wu Zhang (章老师)**, **Jiwen Xia (夏老师)**, 你们是我来到 WesterdijkInstitute遇到的第一群中国人, 您的存在让我在荷兰这个完全陌生的国家有了的归属感! 谢谢你们给予我的教导, 鼓励, 支持, 和陪伴, 让我走过初到荷兰的那段艰难时光。随后**Meihuan (楣缦姐)**, **Xincun Wang (王老师)**, **Ning Jiang (姜宁同学)** 加入了我们这个大家庭, 我们的聚餐活动, 户外活动更加频繁。我会记得我们在, 章老师的公寓, 冬梅姐的住宅, 郊区的大WOK, 市中心的大顺, ah对面的越南店, 等大大小小地点的各种有趣聚餐! 会记得我们的每次说走就走的旅行, 去库肯霍夫公园(阿姆斯特丹)赏郁金香, 去小孩堤防(金德代克)看风车, 去波恩(德国)赏樱花, 等! 还有和亮姐单独的东欧五国的“圣诞音乐之行”。你们在我的荷兰留学生活中留下了浓墨重彩的一笔! 希望各位老师在国内工作顺利, 身体健康! **Lingwei Hou (玲伟师姐)**, 我们相识在疫情期间, 虽然集体的外出活动受到限制, 但是我很享受我们(还有佳丽)一起在所里加班的日子! 20年冬季, 我们时常加班到保安上来催促, 常常卡在最后几分钟离开大楼! 只要师姐拍照的房间亮着, 我就有安全感:)。我也很怀念我们(和佳丽, 佳佳)在K楼110房间的临别聚餐, 我们聊实验, 聊感情, 聊工作, 三个女人一台戏, 那时我们是四个:)。好久没联系了, 真心希望师姐工作顺利! **Yuanyuan Chen (圆圆师姐)**, **Lin Zhao (赵琳)**, **Yanfang Guo (艳芳)**, **Xin Zhang (张鑫)**, 我们相识于后疫情时代:)。特别感谢圆圆师姐在去年年底和今年年初对我找工作所给予的建议和帮助! 怀念坐在fishtank期间和师姐, 赵琳的聊天, 你们太有趣了, 很神奇, 每次和你们聊完就会又充满干劲! 我也很怀念和大家好几次的聚餐(火锅, 做菜, 披萨)聊天, 很温暖, 很放松! 知道一直有人在你身边, 随时可以和你分享喜怒哀乐的感觉太棒了! 真心感谢你们的一路陪伴和支持! 希望师姐回国申请基金顺利, 评职称顺利! 希望赵琳, 艳芳, 张鑫实验顺利, 享受在荷兰的PhD生活!

Special thanks also to my friends and amazing CSC-ers! **顾源, 余小飞**, 我们从研究生就认识, 2018年毕业后, 很幸运和你们一起又再次出国深造。虽然我们在国外联系的不频繁, 但每次与你们聊天都很放松, 很减压。谢谢你们给予的莫大精神支持! **侯玉煌**, 我们真的是有许多共同好友:), 但遗憾的是我们彼此交流并不多。与你正式熟实是在羊角村的游玩, 你刷新了我的认识, 哈哈, 是一个宝藏朋友:) 上次比利时见面太匆忙, 希望有机会可以再次坐下来聊聊天。写这段话时, 其实我在想“要不要和你一起介绍顾源和余小飞认识一下”, 哈哈。**解敏, 张昊师兄, 刘庆午师兄, 吕俊, 王振国**, 因为同在UU读书而有幸认识你们。每次有问题联系你们, 你们积极地回应与帮助都让我觉得很温暖。谢谢**解敏**小姐妹的财产与精神支持, 好几次CSC卡出现问题都是你帮的忙, 给力姐妹! 谢谢**张昊**师兄总是耐心解答我关于选课, UU课程经费的问题。因为一起上写作课意外得知我们竟然是老乡, 莫名有种“他乡遇故知”的熟悉感! 谢谢**刘庆午**师兄组织的春节聚餐并发送“延期申请”文件, 耐心帮助我, 感恩! 谢谢**吕俊**同学分享关于打印的tricks, **王振国**同学分享关于生活的tips, 联系不多, 但是你们每次给的建议都非常中肯。希望大家在海外一切顺利, 回国后保持联系!

Finally, my heartfelt thanks and sincere love to my family in China! 亲爱的爸爸妈妈, 我爱你们! 谢谢你们为所做的, 所付出的一切! 谢谢你们一直支持我做的每一个决定, 并且总是无条件相信我。此外特别感谢你们从来不催婚, 希望你们能继续保持©! 哥哥, 尽管你经常对我不耐烦:(, 我们很难坐下来心平气和的聊天, 但是当家里发生一些重要的事情, 谢谢你一直在! 你的存在让我很安心:)! 我也很爱你! 和爱爸爸妈妈一样爱! 嫂子, 很心水你的性格, 感谢你一直陪在爸爸妈妈哥哥身边, 谢谢你默默的付出! 还有我亲爱的小侄女李奕可(YiCo), 欢迎你的到来, 想到回国马上就要与你见面, 小姑姑很激动, 希望你能开开心心茁壮成长。最后, 我想把我最深的思念和爱意给我在天堂的爷爷, 虽然我们看不见面, 但是我却时常感觉您在我身边~

再过一个月, 我在国外的四年学习就彻底结束了。刚刚过完三十岁生日的我, 回想这四年, 除了提升自己的科研能力之外, 也从身边人身上学习到许多对我个人成长至关重要的



品质：自信和勇敢。不吝于向他人表达自己的赞美与喜爱，也不畏于对事物发表自己的看法与观点！

The End!

“When we reach the end of what we should know, we will be at the beginning of what we should sense.”

— Kahlil Gibran, *Sand and Foam*