



**CHALMERS**  
UNIVERSITY OF TECHNOLOGY

## **Genomic and transcriptomic analysis of the thermophilic lignocellulose-degrading fungus *Thielavia terrestris* LPH172**

Downloaded from: <https://research.chalmers.se>, 2024-04-19 15:36 UTC

Citation for the original published paper (version of record):

Tölgo, M., Hüttner, S., Rugbjerg, P. et al (2021). Genomic and transcriptomic analysis of the thermophilic lignocellulose-degrading fungus *Thielavia terrestris* LPH172. *Biotechnology for Biofuels*, 14(1).  
<http://dx.doi.org/10.1186/s13068-021-01975-1>


N.B. When citing this work, cite the original published paper.

RESEARCH

Open Access



# Genomic and transcriptomic analysis of the thermophilic lignocellulose-degrading fungus *Thielavia terrestris* LPH172

Monika Tölgo<sup>1,2†</sup>, Silvia Hüttner<sup>1,2†</sup>, Peter Rugbjerg<sup>2</sup>, Nguyen Thanh Thuy<sup>3</sup>, Vu Nguyen Thanh<sup>3</sup>, Johan Larsbrink<sup>1,2</sup> and Lisbeth Olsson<sup>1,2\*</sup> 

## Abstract

**Background:** Biomass-degrading enzymes with improved activity and stability can increase substrate saccharification and make biorefineries economically feasible. Filamentous fungi are a rich source of carbohydrate-active enzymes (CAZymes) for biomass degradation. The newly isolated LPH172 strain of the thermophilic Ascomycete *Thielavia terrestris* has been shown to possess high xylanase and cellulase activities and tolerate low pH and high temperatures. Here, we aimed to illuminate the lignocellulose-degrading machinery and novel carbohydrate-active enzymes in LPH172 in detail.

**Results:** We sequenced and analyzed the 36.6-Mb genome and transcriptome of LPH172 during growth on glucose, cellulose, rice straw, and beechwood xylan. 10,128 predicted genes were found in total, which included 411 CAZy domains. Compared to other fungi, auxiliary activity (AA) domains were particularly enriched. A higher GC content was found in coding sequences compared to the overall genome, as well as a high GC3 content, which is hypothesized to contribute to thermophilicity. Primarily auxiliary activity (AA) family 9 lytic polysaccharide monooxygenase (LPMO) and glycoside hydrolase (GH) family 7 glucanase encoding genes were upregulated when LPH172 was cultivated on cellulosic substrates. Conventional hemicellulose encoding genes (GH10, GH11 and various CEs), as well as AA9 LPMOs, were upregulated when LPH172 was cultivated on xylan. The observed co-expression and co-upregulation of genes encoding AA9 LPMOs, other AA CAZymes, and (hemi)cellulases point to a complex and nuanced degradation strategy.

**Conclusions:** Our analysis of the genome and transcriptome of *T. terrestris* LPH172 elucidates the enzyme arsenal that the fungus uses to degrade lignocellulosic substrates. The study provides the basis for future characterization of potential new enzymes for industrial biomass saccharification.

**Keywords:** Biomass degradation, Carbohydrate active enzymes, Cellulose, Filamentous fungi, LPMO, Thermostable enzymes, Transcriptome, Xylan

## Background

The biorefinery concept represents the basis for a more sustainable bio-based economy aimed at converting abundant renewable biomass sources into energy and value-added products. Today, around 40 lignocellulosic biorefineries operate across Europe [1]. Even though lignocellulose is a potential biomass resource, its degradation is impeded by high lignin content and heterogeneity

\*Correspondence: lisbeth.olsson@chalmers.se

†Monika Tölgo and Silvia Hüttner share first authorship

<sup>1</sup> Wallenberg Wood Science Centre, Department of Biology and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden

Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

of its polysaccharide constituents [2, 3]. Biomass saccharification into fermentable monomeric sugars by enzymatic hydrolysis is a crucial step in a biorefinery, but it is hindered by the high cost of enzymes. Indeed, enzymes have been estimated to add 1 USD/gallon to the cost of bioethanol produced from poplar. Thus, there is strong demand for improved enzyme activity and stability [4].

Various potential industrial enzymes exist in nature [5] and the Kingdom Fungi, with more than a million species, represents a particularly rich source [6]. As major biomass degraders, fungi possess a broad array of enzymes suitable for lignocellulose degradation, which are often secreted in large quantities [7]. Thermophilic and thermo-tolerant fungi are especially interesting, as their enzymes can usually endure harsh conditions used in the industry, such as extreme temperatures or pH and harsh solvents [8, 9]. For example, biomass hydrolysis by the industrial *Trichoderma reesei* enzymes in a separate hydrolysis–fermentation process (SHF), is performed at 45–50 °C and pH 5 and therefore additional enzymes that are added to this process to enhance hydrolysis further should show high activity under the same conditions. Thermostable enzymes can lower industrial processing costs as they can achieve faster reaction rates, greater stability, and are more easily adjustable to various setups [10]. Yet, it should be kept in mind that novel enzymes from thermophiles are not necessarily thermostable when produced heterologously [11].

*Thielavia terrestris* (syn *Thermothielavioides terrestris* [12]) is a well-known filamentous fungus identified in 1983 as a potential source of thermostable industrial enzymes based on successful (hemi)cellulase assays [13, 14]. The species is a thermophilic saprobic Ascomycete isolated mainly from soil and compost in Asia [15–17] and from a cave cricket species in North America [18]. *T. terrestris* also played a pivotal role in the discovery of the cellulase-boosting effect of the glycoside hydrolase family 61 (GH61) proteins [19–23], today known as auxiliary activity family 9 (AA9) lytic polysaccharide monooxygenases (LPMOs) [24]. As described by Merino and Cherry [19], cultivation broth from *T. terrestris* primed for cellulase production showed striking synergy in degrading pre-treated corn stover when supplemented with the enzyme cocktail Celluclast. In 2011, *T. terrestris* strain NRRL 8126 and *Myceliophthora thermophila* ATCC 42464 were the first thermophiles whose genomes were fully sequenced and the first filamentous fungi with known telomere-to-telomere sequences [25]. The same study showed that *T. terrestris* could potentially degrade all plant cell wall polysaccharides and the fungus hydrolyzed alfalfa straw at temperature optima of 40 °C and 60 °C. As shown by proteomics analyses [26] and detailed biochemical characterization [15, 16, 20, 27–33],

*T. terrestris* produces an array of biomass-degrading enzymes. However, no study has elucidated the gene expression mechanisms underlying the lignocellulolytic machinery of the fungus in detail.

In recent years, it has become clear that genetic or gene expression differences between fungal strains of the same species are not uncommon [17, 34–37]. In this study, we set out to sequence and analyze the genome and transcriptome of the newly isolated *T. terrestris* strain LPH172, which is characterized by superior enzymatic activity, thermostability, and pH stability [17]. Our current study aimed to elucidate the lignocellulose-degrading machinery of the fungus in detail and identify novel carbohydrate-active enzymes (CAZymes). We observed some genomic differences between LPH172 and the previously sequenced strain NRRL 8126. To corroborate genomic CAZyme analysis with transcriptome data, we grew the fungus on four substrates: glucose, Avicel, rice straw, and beechwood xylan. We observed that the fungus relied mainly on LPMOs and canonical cellulases when grown on cellulosic substrates and on hemicellulosic substrates, more conventional hemicellulases were induced together with some LPMOs. Interestingly, we also report co-expression and co-upregulation between LPMOs and other AA enzymes.

## Results

### Strain identification

We previously isolated the *T. terrestris* strain LPH172 from compost in Northern Vietnam and showed that it could be exploited as an industrially relevant enzyme producer [17]. To confirm the identity of the fungus, first, we used two common fungal housekeeping genes [38] encoding transcription elongation factor 1- $\alpha$  and  $\beta$ -tubulin. The homologous gene sequences used for the identification procedure are listed in Additional file 1. Both housekeeping genes were 100% identical (e-value 0) to the *T. terrestris* NRRL 8126 homologues.

Furthermore, we used phylogenetic analysis of closely related species to confirm the strain identity. Three maximum likelihood phylogenetic trees were constructed based on the internal transcribed spacer (*ITS-D1D2*), RNA polymerase II (*RPB2*) and  $\beta$ -tubulin (*TUB2*) genes. In all three constructed trees, the *T. terrestris* strain LPH172 clustered closely together with four other *T. terrestris* strains (Additional file 1: Figure S1).

Both of these results confirmed the fungus in the current study to be a strain of *T. terrestris*.

### Growth on different carbohydrates

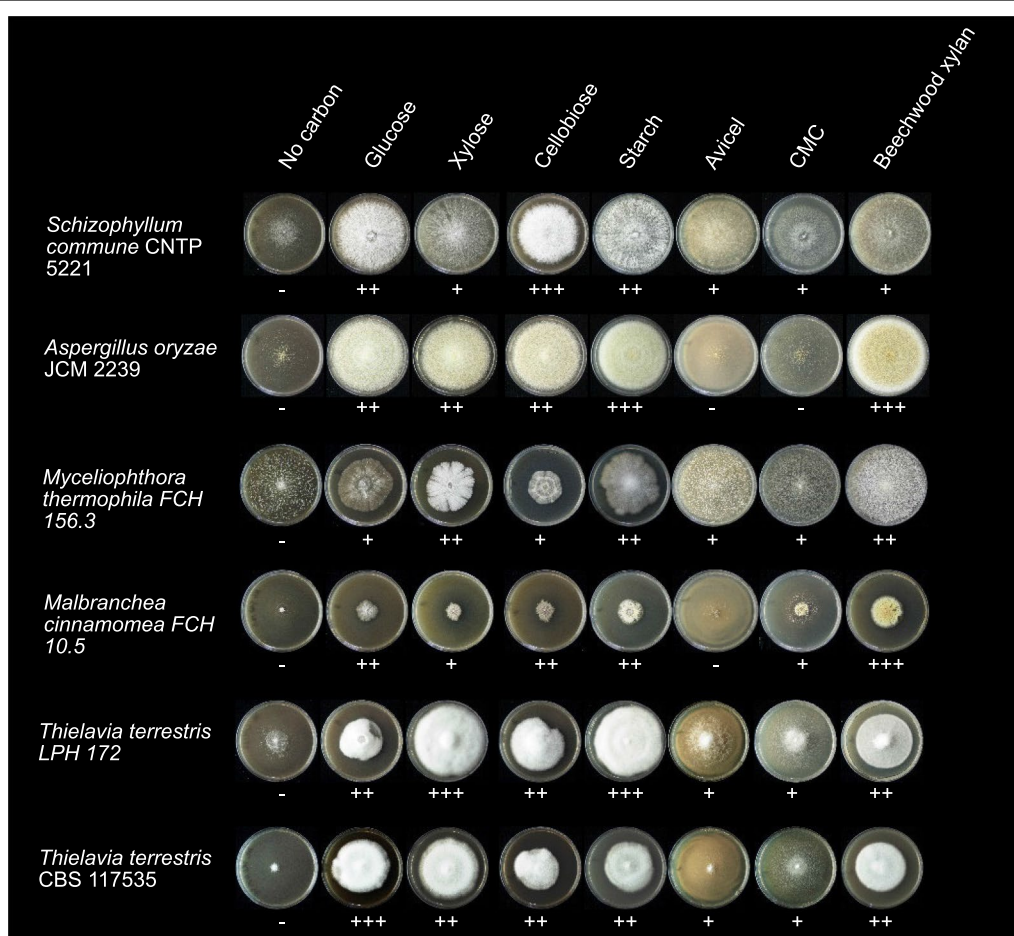
To assess the ability of *T. terrestris* LPH172 to utilize different carbon sources, the strain was grown on various defined substrates on agar. Growth was measured by the

diameter and density of mycelia and was compared to a selection of known mesophilic and thermophilic biomass degraders (Fig. 1). *T. terrestris* LPH172 grew best on starch and xylose, followed by glucose, cellobiose, and beechwood xylan, whereas only modest growth was observed on the cellulosic substrates Avicel and carboxymethyl cellulose (CMC). This finding suggests relatively high activity of amylases, xylanases, and  $\beta$ -glucosidases. Direct comparison to the previously sequenced *T. terrestris* CBS 117535 (GenBank acc. nr PRJNA249224) showed that LPH172 grew slightly better on most substrates except glucose. Good growth was observed on pectin and inulin (a fructose-based polymer), whereas growth on locust bean gum and guar gum (galactomannans), as well as bark powder was poor (Additional file 2).

## Genome characterization

### General features

The genome of *T. terrestris* LPH172 was sequenced on a PacBio RS II instrument by GATC Biotech (Constance, Germany); it yielded 5,27,523 reads comprising over 7 billion bases. Table 1 gives an overview of the sequenced *T. terrestris* LPH172 genome. Its size was determined to be 36.6 Mb and guanine and cytosine (GC) content was 54.80%. Assembly quality, based on basic sequence statistics, was high as revealed by an average contig size (N50) of 3 Mb and N50 read length of 19,832. To assess completeness and integrity of the genome assembly, Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis was performed [39]. Over 98% of BUSCO genes in the LPH172 genome were complete, indicating excellent assembly integrity. Gene prediction algorithms identified 10,128 protein-coding genes.



**Fig. 1** Growth of *T. terrestris* LPH172 and other biomass-degrading filamentous fungi on different carbon sources. Seven different carbohydrate substrates at 2% (w/v) were used as sole carbon sources for growth on agar plates: monosaccharides (glucose, xylose), disaccharides (cellobiose), and polysaccharides (starch, Avicel, carboxymethyl cellulose—CMC, beechwood xylan). No carbon source was added in the control. The plates were incubated at 30 °C (*S. commune*, *A. oryzae*) or 50 °C (*M. thermophila*, *M. cinnamomea*, *T. terrestris*) for 2–7 days. Growth was evaluated visually and categorized from – (no growth) to +++ (very good growth), depending on the extent and density of the mycelium

**Table 1** Overview of *T. terrestris* LPH172 genome as sequenced in this study

Genome assembly	
Number of nucleotides	36,579,697
GC content	54.80%
N50 (bp)	3,006,457
Number of protein-coding genes	10,128
Average gene length (bp)	1628
Average CDS length (bp)	1355
Average number of exons per gene	3
Average exon length (bp)	460
Longest genes (bp)	25,297
Longest exons (bp)	9745
Shortest genes (bp)	39
Shortest exons (bp)	3
Fraction of genome covered by	
Genes	45.10%
Exons	37.80%
Introns	7.30%
Gene annotation	
Genes with functional annotation	8879
Genes without functional annotation	1249
Genes annotated (BLASTp, e-value < 1e <sup>-6</sup> )	6114

The size of fungal genomes can vary by orders of magnitude and the average for Ascomycota is 36.91 Mb [40, 41]. Table 2 gives a brief overview of the characteristics of the LPH172 genome compared to other industrial and lignocellulose-degrading fungi with varied origin and thermostability. Even though the genome of *T. terrestris* NRRL 8126 (GenBank assembly nr GCA\_000226115), sequenced in 2010, was slightly larger than that of LPH172, our analysis suggested LPH172 contained approximately 200 more genes. This discrepancy, in

addition to inherent differences between the two strains, is likely a consequence of ongoing improvements in sequencing and annotation. The genome size of LPH172 was similar to those of other fungi listed in Table 2, as well as to the average Ascomycota genome. The same was true for the average genome GC content. The average gene length in strain LPH172 was 1628 bp and the average coding sequence was 1355 bp (Table 1). On average, three exons per gene were predicted, with exons covering 45.10% of the genome. 88% of the genes could be functionally annotated with BLASTp, 69% of which with high certainty (e-value < 1e<sup>-6</sup>).

### Thermostability features

Although there is no clear consensus on the causes contributing to elevated optimum growth temperatures and thermotolerance in fungi, possible factors include a reduction in genome size [49], higher average GC content in coding regions, and greater GC content in the third position of codons (GC3 content) [25, 50].

In contrast to the thermophilic Ascomycete *Malbranchea cinnamomea* [43], the genome of *T. terrestris* LPH172 was not smaller compared to that of other mesophilic fungi (Table 2). Normalized (gene length) GC content in gene-coding sequences was 63.5%, which was higher than the genome average of 54.80%. When looking only at the subset of genes encoding CAZymes, the average normalized GC content was even higher (64.5%). Normalized GC3 content in LPH172 was also high, amounting to 80.7% in coding sequences and 85.7% in CAZyme-encoding sequences. We also detected gene TT\_05393, encoding an unknown protein with 33% identity (e-value 1.3e<sup>-19</sup>) to the known thermotolerance gene *THTA* from *Aspergillus fumigatus* (GenBank: AY560012.1) [51].

**Table 2** Genome characteristics in *T. terrestris* LPH172 and different industrial and lignocellulose-degrading fungi

Organism	Phylum	Thermostolerance	Genome (Mb)	Genome GC (%)	Number of protein-coding genes	Source
<i>A. oryzae</i>	Ascomycota	Mesophilic	37.9	47.2	12,030	[42]
<i>M. thermophila</i>		Thermophilic	38.7	51.4	9110	[25]
<i>M. cinnamomea</i>		Thermophilic	25.0	49.8	9437	[43]
<i>P. anserina</i>		Mesophilic	35.0	52.0	10,588	[44]
<i>T. terrestris</i> LPH172		Thermophilic	36.6	54.8	10,128	This article
<i>T. terrestris</i> NRRL 8126	Basidiomycota	Thermophilic	36.9	54.7	9813	[25]
<i>G. trabeum</i>		Mesophilic	37.2	52.9	11,755	[45]
<i>S. commune</i>		Mesophilic	38.5	N/A	13,210*	[46]
<i>R. pusillus</i>	Zygomycota	Thermophilic	25.6	45.0	10,898	[47]
<i>R. oryzae</i>		Mesophilic	39.1	35.4	13,895	[48]

\*nr of genes

### CAZyme comparison with other fungi

Plant biomass-degrading and other CAZymes are catalogued into classes, families and subfamilies in the Carbohydrate Active enZymes (CAZy) database (<http://www.cazy.org/>) [52]. The number of individual CAZyme domains and distribution across different CAZy families in *T. terrestris* LPH172 was analyzed and compared to other known fungal biomass degraders to assess the propensity for lignocellulose degradation (Table 3). In total, 411 individual CAZy domains were detected in LPH172 using dbCAN2 (HMMER algorithm) (Additional file 3). Most CAZy domains were found to be GHs (201 candidates), with GH18 ( $n = 15$ ), GH16 ( $n = 14$ ), GH3 ( $n = 12$ ), and GH47 ( $n = 10$ ) being the most abundant families. There were also 86 glycosyl transferase (GT) domains, 4 polysaccharide lyase (PL) domains, 26 carbohydrate esterase (CE) domains, 83 AA domains, and 11 carbohydrate-binding module (CBM) domains. Compared to strain NRRL 8126, two more GH (one GH16 and one GH47) domains were identified in LPH172, as well as one additional AA12, one GT2, and one CE1 domain (Additional file 3). *T. terrestris* LPH172 had a relatively low number of PL domains compared to other fungi (Fig. 2), but a larger complement of AA family domains, particularly AA7 ( $n = 20$ ), AA9 ( $n = 18$ ) and AA3 ( $n = 16$ ) (Fig. 3). Five members of AA11 (chitin-cleaving) LPMO domains were detected in both *T. terrestris* strains, but no AA13 (starch-cleaving LPMOs) or AA14 (xylan-cleaving LPMOs) domains were observed. LPH172 and NRRL 8126 were the only fungi, among the ones selected, presenting an AA16, a recently characterized C1-hydroxylating LPMO [53].

Putative candidates for CAZymes capable of degrading all major lignocellulosic polymers (cellulose, xylan, xyloglucan, (galacto)glucoman, pectin, and lignin), as

well as starch, inulin, and chitin were found. This finding was in line with growth of *T. terrestris* on all these carbon sources (Additional file 2). At least ten putative homologues of known transcription factors directly regulating (hemi)cellulose utilization in ascomycetous fungi [54] were also detected in the genome of *T. terrestris* LPH172 (Additional file 4).

### Transcriptome analysis

**Highly expressed CAZyme-encoding genes during growth on Avicel, rice straw, and beechwood xylan** To verify the genome annotation of LPH172 and analyze gene expression, in particular of CAZyme-encoding genes, the transcriptome was analyzed under different growth conditions. The fungus was grown in shake flasks on four substrates—glucose, Avicel, rice straw and beechwood xylan—and total mRNA was extracted and sequenced. Glucose was chosen as reference monosaccharide because its degradation involves a limited number of CAZymes and should, therefore, reflect expression of mostly constitutive CAZyme-encoding genes. Beechwood xylan, comprising a xylan backbone with 4-*O*-methyl glucuronic acid side groups, was selected to detect CAZymes required for hardwood hemicellulose degradation [43]. Rice straw, which contains approximately 12% lignin, 38% cellulose, and 25% hemicellulose [55] was chosen to represent a complex, heterogeneous substrate requiring a large array of different CAZymes for degradation. Importantly, rice straw has also vast potential as feedstock in biorefinery applications. Finally, Avicel, which is up to 98% cellulose [56, 57], was selected to identify enzymes required to degrade a highly crystalline and recalcitrant cellulosic substrate. Initially, our transcriptome analysis also included corn cob xylan as an arabinoxylan-containing substrate model for cereals. The results were not included

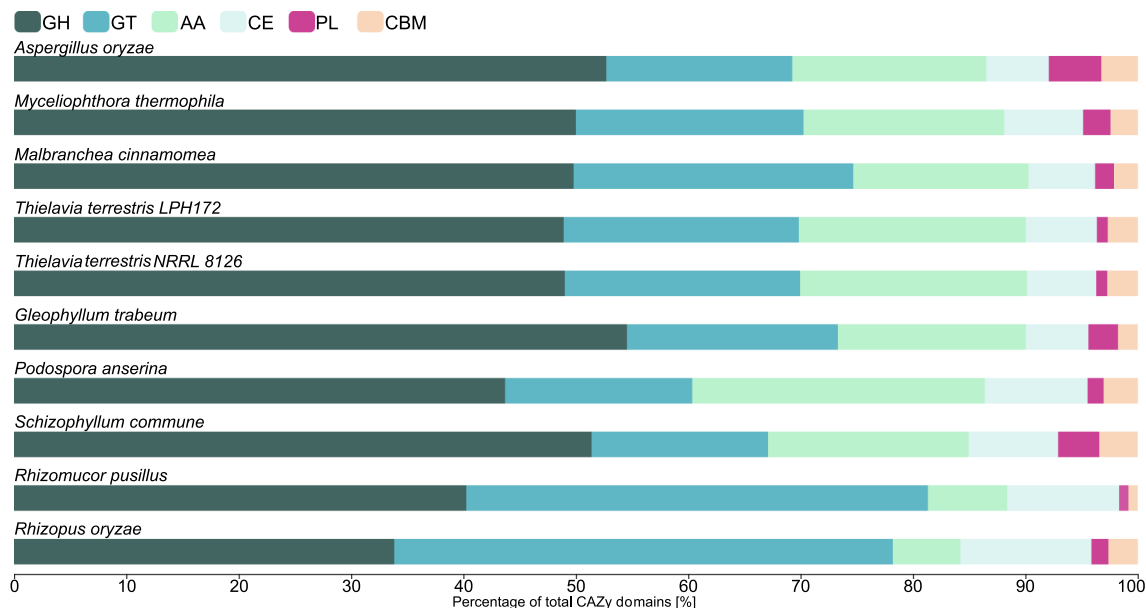
**Table 3** Comparison of the number of individual CAZy domains in *T. terrestris* LPH172 and other filamentous fungi

	GH	GT	PL	CE	AA	CBM	Total
<i>Aspergillus oryzae</i>	292	92	26	31	96	18	555
<i>Myceliophthora thermophila</i>	185	75	9	26	66	9	370
<i>Malbranchea cinnamomea</i>	118	59	4	14	37	5	237
<i>Thielavia terrestris</i> LPH172	201	86	4	26	83	11	411
<i>Thielavia terrestris</i> NRRL 8126	199	85	4	25	82	11	406
<i>Gloeophyllum trabeum</i>	186	64	9	19	57	6	341
<i>Podospora anserina</i>	215	82	7	45	128	15	492
<i>Schizophyllum commune</i>	239	73	17	37	83	16	465
<i>Rhizomucor pusillus</i>	97	99	2	24	17	2	241
<i>Rhizopus oryzae</i>	90	118	4	31	16	7	266

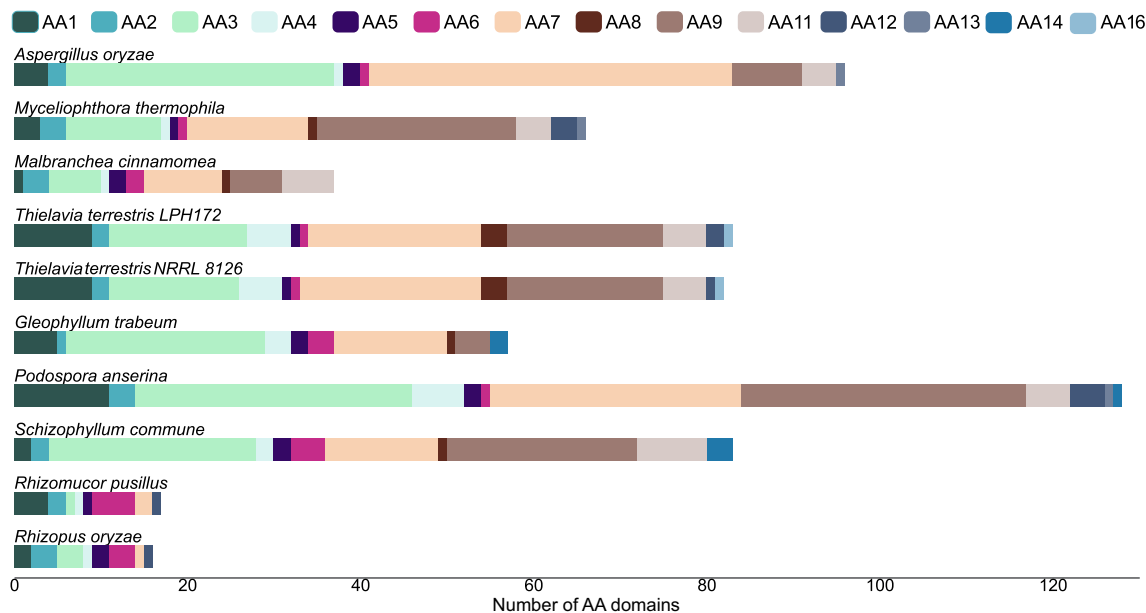
All CAZy domains were identified using dbCAN2 (HMMER algorithm)

CE10 family domains were excluded

GH glycoside hydrolase, GT glycoside transferase, AA auxiliary activity, CE carbohydrate esterase, PL polysaccharide lyase, CBM carbohydrate-binding module



**Fig. 2** Relative numbers of CAZy domains from six CAZy families in various filamentous fungi. For each species, the numbers of predicted CAZyme domains were normalized to the total number of predicted CAZyme domains. GH, glycoside hydrolase; GT, glycoside transferase; AA, auxiliary activity; CE, carbohydrate esterase; PL, polysaccharide lyase; CBM, carbohydrate-binding module. Predictions were made with dbCAN2 (HMMER algorithm). CE10 family domains were excluded



**Fig. 3** Number of auxiliary activity (AA) CAZyme family domains in different filamentous fungi. For each species, the number of predicted AA domains categorized into AA families 1–16 is shown. Predictions were made with dbCAN2 (HMMER algorithm)

because the purchased corn cob xylan turned out to be composed only of xylo-oligomers [58]. Transcriptome data from RNAseq was used to refine gene annotation

through *ab initio* training with GeneMark v4.3 and an evidence-guided build with MAKER package v3.01.1. Results are summarized in Tables 4–6.

To identify which CAZyme-encoding genes were the most highly expressed (by transcript abundance) on each chosen substrate, we looked at the top 20 (arbitrary number) candidates under each growth condition, ranked by their average transcripts per million (TPM) value. The complete list of all expressed genes with their TPM values is available in Additional file 5.

Among the top 20 highly expressed CAZyme-encoding genes on Avicel (Table 4) there were in total 11 putative cellulase and LPMO-encoding genes—six cellulases and five LPMOs. These highly expressed cellulase genes included one encoding a putative GH7-CBM1 exoglucanase (TT\_05546), a GH6 cellobiohydrolase (TT\_06655) and four endoglucanases GH7-CBM1 (TT\_05797), GH5\_5 (TT\_01019 and TT\_00224) and GH45 (TT\_09000). All five highly expressed LPMO-encoding genes (TT\_07456, TT\_08370, TT\_04350, TT\_06407, TT\_01736) belonged to the AA9 family. A gene encoding a CAZyme possibly acting in synergy with LPMOs was also highly expressed—a putative AA3-AA8 cellobiose dehydrogenase encoding gene (TT\_04380). Genes encoding five putative xylan-active enzyme genes were also highly abundant on Avicel, encoding a CE5-CBM1 acetylxylan esterase (TT\_08166), a feruloyl esterase with

a CBM1 domain (TT\_06499), two GH11 endo- $\beta$ -1,4-xylanases (TT\_03075 and TT\_01839) and a GH10-CBM1 endo- $\beta$ -1,4-xylanase (TT\_08161). In general, seven of the 20 highly expressed CAZyme genes on Avicel also included a cellulose-binding CBM1 domain.

When the fungus was cultivated on rice straw (Table 5), only one putative cellulase gene (GH131 endoglucanase; TT\_01081) was among the top 20 most highly expressed CAZyme-encoding genes. However, five AA9 LPMOs were again highly expressed (TT\_07456, TT\_08370, TT\_04352, TT\_06407, TT\_03770) - of which TT\_07456, TT\_08370 and TT\_06407 were among the top 20 highly expressed CAZyme genes also on Avicel. None of these putative LPMO-encoding genes expressed during growth on rice straw included a CBM1. For possible xylan backbone degradation, a putative GH11 endo- $\beta$ -1,4-xylanase (TT\_01839) encoding gene was highly expressed and a putative GH11-CBM1 endo- $\beta$ -1,4-xylanase (TT\_03075) as well. Both were also highly expressed on Avicel. In addition, a gene encoding a putative CE16 acetyl esterase (TT\_06012) possibly cleaving acetyl groups from xylan was among the highly expressed CAZyme genes. Interestingly, on rice straw, genes encoding putative mannosyltransferases belonging to the CAZy families GT15 (TT\_07992 and TT\_03483), GT32 (TT\_08079) and GT2 (TT\_05489) were also among the top 20 most highly expressed CAZyme-encoding genes. Moreover, a CE5-CBM1 cutinase (TT\_08797) was among these top 20 expressed CAZyme genes on rice straw, possibly active on plant cuticle.

When grown on beechwood xylan, curiously, only two putative xylanase encoding genes were among the 20 most highly expressed genes coding for CAZymes (Table 6). These encoded a putative CE5-CBM1 acetylxylan esterase (TT\_08166) and a GH11 endo- $\beta$ -1,4-xylanase (TT\_01839), the last one also highly expressed on Avicel and rice straw. No AA9 LPMO-encoding genes were among the 20 most highly expressed CAZyme genes on beechwood xylan, though, a gene encoding a putative AA11 LPMO (TT\_09754) was among the top 20 highly expressed genes. Interestingly, several genes encoding putative transferase genes were also highly expressed on beechwood xylan, like on rice straw. These included two genes encoding putative GH72  $\beta$ -1,3-glucanosyltransferases (TT\_03756 and TT\_03035) and a gene encoding a putative GT32 mannosyltransferase. Genes coding for putative GH16 glucosidases were also among the top 20 most abundant transcripts on beechwood xylan (TT\_08846, TT\_09509 and TT\_02410). Interestingly, also genes encoding a putative GH37 trehalase (TT\_00335) and a putative AA2 catalase-peroxidase (TT\_08899) were highly expressed on this hemicellulosic substrate. During growth on both rice straw and

**Table 4** Twenty most highly expressed CAZyme-encoding genes during growth of *T. terrestris* LPH172 on Avicel

Transcript ID	TPM	CAZy domain(s)	Putative function
TT_05546	12001	GH7-CBM1	Exoglucanase
TT_07456	8862	AA9	LPMO
TT_06655	8643	GH6	1,4- $\beta$ -D-Glucan cellobiohydrolase
TT_08370	8271	AA9	LPMO
TT_06499	6276	CBM1	Feruloyl esterase
TT_00224	4790	GH5_5	Endoglucanase
TT_05797	4654	GH7-CBM1	Endoglucanase
TT_03075	4576	GH11-CBM1	Endo-1,4- $\beta$ -xylanase
TT_04350	4537	AA9-CBM1	LPMO
TT_07210	3659	GH30_7	Glucosylceramidase
TT_08166	3512	CE5-CBM1	Acetylxylan esterase
TT_06407	3218	AA9	LPMO
TT_09000	3188	GH45	Endoglucanase
TT_01736	2677	AA9-CBM1	LPMO
TT_01019	2322	GH5_5	Endoglucanase
TT_04380	1599	AA3_1-AA8	Cellobiose dehydrogenase
TT_01839	1429	GH11	Endo-1,4- $\beta$ -xylanase
TT_08161	1335	GH10-CBM1	Endo-1,4- $\beta$ -xylanase
TT_03756	1104	GH72	1,3- $\beta$ -Glucanosyltransferase
TT_01913	919	GH132	Secreted $\beta$ -glucosidase

TPM values indicate average normalized transcript levels from three biological replicates. CAZy domains were predicted by dbCAN2 and functions were annotated by BLASTp search against the UniProt/Swiss-Prot reference dataset

**Table 5** Twenty most highly expressed CAZyme-encoding genes during growth of *T. terrestris* LPH172 on rice straw

Transcript ID	TPM	CAZy domain(s)	Putative function
TT_01839	8312	GH11	Endo-1,4- $\beta$ -xylanase
TT_07456	963	AA9	LPMO
TT_08370	462	AA9	LPMO
TT_07992	418	GT15	Glycolipid 2- $\alpha$ -mannosyltransferase
TT_08797	400	CE5	Cutinase
TT_08097	276	GT32	Initiation-specific $\alpha$ -1,6-mannosyltransferase
TT_05010	249	GH25	<i>N,O</i> -Diacetylmuramidase
TT_01081	241	GH131	Endoglucanase
TT_04352	234	AA9	LPMO
TT_05489	178	GT2	Dolichol-phosphate mannosyltransferase subunit
TT_03075	171	GH11-CBM1	Endo-1,4- $\beta$ -xylanase
TT_06407	167	AA9	LPMO
TT_08740	166	CBM52	Uncharacterized secreted protein
TT_09851	162	GH128	Alkali-sensitive linkage protein
TT_03483	158	GT15	Mannosyltransferase
TT_05929	139	GT31	$\beta$ -1,3- <i>N</i> -Acetylglucosaminyltransferase
TT_03756	135	GH72	1,3- $\beta$ -Glucanoyltransferase
TT_06012	125	CE16	Acetylesterase
TT_01913	119	GH132	Secreted $\beta$ -glucosidase
TT_03770	117	AA9	LPMO

TPM values indicate average normalized transcript levels from three biological replicates. CAZy domains were predicted by dbCAN2 and functions were annotated by BLASTp search against the UniProt/Swiss-Prot reference dataset

**Table 6** Twenty most highly expressed CAZyme-encoding genes during growth of *T. terrestris* LPH172 on beechwood xylan

Transcript ID	TPM	CAZy domain(s)	Putative function
TT_05010	2452	GH25	<i>N,O</i> -Diacetylmuramidase
TT_01913	1466	GH132	Secreted $\beta$ -glucosidase
TT_03756	1366	GH72	1,3- $\beta$ -Glucanoyltransferase
TT_09441	1359	CE9	<i>N</i> -Acetylglucosamine-6-phosphate deacetylase
TT_08166	1329	CE5-CBM1	Acetylxylen esterase
TT_08780	903	GH17	Glucan endo-1,3- $\beta$ -glucosidase
TT_03035	892	GH72	1,3- $\beta$ -Glucanoyltransferase
TT_08846	743	GH16-CBM18	Glycosidase
TT_09754	736	AA11	LPMO
TT_01404	608	GH17	$\beta$ -Glucosidase
TT_00335	516	GH37	Trehalase
TT_09509	487	GH16	Glycosidase
TT_06210	471	GH30_3	Endo-1,6- $\beta$ -D-glucanase
TT_08740	469	CBM52	Uncharacterized secreted protein
TT_00639	459	GH20	$\beta$ -Hexosaminidase
TT_01839	419	GH11	Endo-1,4- $\beta$ -xylanase
TT_08097	383	GT32	Initiation-specific $\alpha$ -1,6-mannosyltransferase
TT_08899	375	AA2	Catalase-peroxidase
TT_05769	317	GH128	Alkali-sensitive linkage protein
TT_02410	310	GH16	Extracellular glycosidase

TPM values indicate average normalized transcript levels from three biological replicates. CAZy domains were predicted by dbCAN2 and functions were annotated by BLASTp search against the UniProt/Swiss-Prot reference dataset

beechwood xylan, a gene encoding a putative CAZyme with a CBM52 domain was highly expressed (TT\_08740). A putative GH72  $\beta$ -1,3-glucanotransferase TT\_03756 was also among the most highly expressed genes on all three substrates.

**Upregulated CAZyme-encoding genes on Avicel, rice straw, and beechwood xylan** Absolute gene expression levels do not show the full spectrum of available lignocellulose-degrading enzymes in the organism, because many enzymes are sufficiently active at low concentration and some are constitutively expressed. Therefore, to understand which genes were induced under the tested conditions (Avicel, rice straw or beechwood xylan), we investigated the differential expression of CAZyme-encoding genes with respect to glucose as reference using edgeR [59]. We focused on the transcripts that were significantly highly expressed (i.e., upregulated) compared to glucose in the test conditions ( $p \leq 0.05$ ) (Additional file 6). A combined list of the top 40 most highly upregulated CAZyme-encoding genes on all three substrates is shown in Fig. 4.

On Avicel, the most highly upregulated genes encoded a putative AA9-CBM1 LPMO gene, (TT\_01736, log<sub>2</sub> fold-change (log<sub>2</sub>FC) 13). Other putative AA9 LPMO-encoding genes were also highly upregulated: TT08370 and TT07456 had log<sub>2</sub>FC 11 and TT04350, TT07962 and TT6407 all had log<sub>2</sub>FC 10. A gene encoding a putative AA3-AA8 cellobiose dehydrogenase was also among the most highly upregulated (TT04380, log<sub>2</sub>FC 9). All AA9 (besides TT07962) and the AA3-AA8 CAZyme-encoding genes presented also very high TPM values, indicating both high upregulation and high expression levels. Other putative CAZyme genes highly upregulated on Avicel encoded mainly GH7 (TT05797 with log<sub>2</sub>FC 8, TT05512 with log<sub>2</sub>FC 12, TT03032 with log<sub>2</sub>FC 6, TT05546 with log<sub>2</sub>FC 8), GH5 (TT01019 with log<sub>2</sub>FC 6, TT03238 with log<sub>2</sub>FC 4, TT00224 with log<sub>2</sub>FC 9) and GH45 (TT09000 with log<sub>2</sub>FC 5) endo- and exo-glucanases. One putative GH6 cellobiohydrolase encoding gene was also highly upregulated (TT06655, log<sub>2</sub>FC 8). Interestingly,

numerous non-cellulose acting CAZyme-encoding genes were also upregulated on Avicel. These included two feruloyl esterases (TT\_06499 with log<sub>2</sub>FC 9 and TT\_09709 with log<sub>2</sub>FC 5), five GH10 and GH11 xylanases (TT\_06197 with log<sub>2</sub>FC 9, TT\_03205 with log<sub>2</sub>FC 5, TT\_03075 with log<sub>2</sub>FC 7, TT\_01839 with log<sub>2</sub>FC 7, TT\_08161 with log<sub>2</sub>FC 6), a CE1 acetylxylan esterase (TT\_05636 with log<sub>2</sub>FC 8), a CE12 rhamnogalacturonan acetyltransferase (TT\_02368 with log<sub>2</sub>FC 6), a CE16 acetyltransferase (TT\_06012 with log<sub>2</sub>FC 7) and CE5-CBM1 acetylxylan esterase (TT\_08166 with log<sub>2</sub>FC 6). Two putative AA7 oxidoreductase encoding genes were highly upregulated on Avicel (TT\_03025 with log<sub>2</sub>FC 8 and TT\_06681 with log<sub>2</sub>FC 4). Two GH5\_7 mannan endo- $\beta$ -1,4-mannosidase (TT09640 and TT06537) encoding genes were also greatly upregulated on this substrate.

On rice straw, the most highly upregulated genes encoded a putative CE1 acetylxylan esterase (TT\_05636 with log<sub>2</sub>FC 16), an AA9 LPMO (TT\_09068 with log<sub>2</sub>FC 13) and a GH11 endo- $\beta$ -1,4-xylanase (TT\_01839 with log<sub>2</sub>FC 12). In general, the set of forty most highly upregulated putative CAZyme-encoding genes on rice straw shared some candidates with Avicel, such as nine AA9 LPMO genes and four hemicellulose-active GH10 and GH11 xylanases, a GH7 endo- $\beta$ -1,4-glucanase and CE1, CE5 and CE16 esterases. A putative gene encoding a CBM50 domain (TT\_00910) was highly upregulated both on Avicel and rice straw. In addition to the AA9 LPMO genes highly upregulated on Avicel as well, on rice straw two more AA9 genes appeared: TT\_03770 and TT\_07455 (both with log<sub>2</sub>FC 6). Generally, more hemicellulose-active enzyme encoding genes were among the forty highly upregulated genes on rice straw than on Avicel, which is expected with the more complex substrate composition of rice straw. These genes encoded putative GH10 (TT\_09033 with log<sub>2</sub>FC 8) and GH11 (TT\_02489 with log<sub>2</sub>FC 8) endo- $\beta$ -1,4-xylanases, a GH62  $\alpha$ -L-arabinofuranosidase (TT\_09005 with log<sub>2</sub>FC 10), two CE3 esterases (TT\_07717 and TT\_08970, both with log<sub>2</sub>FC 6), a CE1 feruloyl esterase (TT\_06092 with

(See figure on next page.)

**Fig. 4** Combined expression and upregulation values of CAZyme-encoding genes during cultivation on four different substrates. Putative CAZyme-encoding genes involved in biomass degradation were analyzed for their expression levels (TPM, transcripts per million) from three biological replicates, as well as their differential expression (log<sub>2</sub>FC). The heatmap shows a combination of top forty most highly upregulated CAZyme-encoding genes on three substrates Avicel, rice straw (RS), beechwood xylan (BX) when compared to glucose (Glc). Shading ranges from low expression (light blue) to high expression (magenta). Log<sub>2</sub> fold-change (log<sub>2</sub>FC) shows gene expression during cultivation on Avicel, RS, and BX compared to cultivation on glucose. Shading of upregulated genes (i.e., gene transcripts more abundant on Avicel, RS, and/or BX than on glucose) ranges from light yellow (low upregulation) to dark green (high upregulation). Downregulated genes or genes for which no differential expression was detected or where upregulation was not significant are indicated by blank cells. Only significantly upregulated genes are shown ( $p \leq 0.05$ ). All numbers were rounded to the nearest integer. The putative activities of the gene products are based on BLASTp predictions. CAZy domains were analyzed with dbCAN2. The presence of putative signal peptides (SP), predicted by SignalP 4.0, is indicated by a small s. Putative substrates of the CAZymes are: C, cellulose; Ch, chitin; GM, glucomannan; L, lignin; P, pectin; S, starch; X, xylan; XG, xyloglucan

Transcript ID	Predicted activity	SP	CAZy-domain(s)	TPM				log2FC			Putative substrate(s)
				Avicel	RS	BX	Glc	Avicel	RS	BX	
TT 00150	Laccase		AA1	6	22	56	7		4	3	L
TT 07374	Multicopper oxidase		AA1 2	9	0	14	3	2		2	L
TT 09261	Laccase	s	AA1 3	3	0	0	0	7	2	3	L
TT 05531	Catalase peroxidase	s	AA2	10	0	62	15			2	
TT 04380	Cellobiose dehydrogenase	s	AA3 1-AA8	1599	4	4	2	9	3		C
TT 08234	Oxygen-dependent choline dehydrogenase	s	AA3 2	235	2	108	9	5		4	C
TT 03025	FAD-linked oxidoreductase	s	AA7	1	0	0	0	8	5		C
TT 06681	Oxidoreductase	s	AA7	29	8	2	2	4	4		C
TT 02325	Cytochrome-domain containing protein	s	AA8	3	61	0	0	3	10		C
TT 09190	Cytochrome-domain containing protein	s	AA8	7	30	1	7		5		C
TT 01683	LPMO		AA9	91	6	4	3	5	3		C
TT 01736	LPMO	s	AA9-CBM1	2677	30	20	0	13	9	6	C
TT 03770	LPMO		AA9	41	117	4	10	2	6		C
TT 04350	LPMO	s	AA9-CBM1	4537	9	13	6	10	3	1	C
TT 04352	LPMO	s	AA9	181	234	3	5	5	8		C
TT 05592	LPMO		AA9	11	0	7	1	4	1	3	C
TT 06268	LPMO	s	AA9	7	47	4	0	6	11	5	C
TT 06407	LPMO		AA9	3218	167	32	4	10	8	3	C
TT 07455	LPMO		AA9	8	7	0	0	4	6		C
TT 07456	LPMO		AA9	8862	963	9	4	11	10	1	C
TT 07962	LPMO		AA9	60	14	1	0	10	10	3	C
TT 08370	LPMO	s	AA9	8271	462	21	3	11	10	3	C
TT 09068	LPMO		AA9	0	9	0	0	6	13	3	C
TT 09080	LPMO	s	AA9	1	27	1	1	1	8		C
TT 09453	PQQ-dependent pyranose dehydrogenase		AA12	31	1	120	4	3		5	
TT 05636	Acetyl xylan esterase	s	CE1	2	64	0	0	8	16		X
TT 06092	Feruloyl esterase	s	CE1	10	91	6	2	2	8	2	X
TT 02368	Rhamnogalacturonan acetyl esterase	s	CE12	2	0	1	0	6	6	4	P
TT 06012	Acetyl esterase	s	CE16	915	125	8	8	7	6		X
TT 07717	Multidomain esterase	s	CE3	5	84	1	6		6		X
TT 08970	Multidomain esterase	s	CE3	3	46	1	4		6		X
TT 05762	Acetyl xylan esterase	s	CE5	17	16	44	1	5	7	6	X
TT 08166	Acetyl xylan esterase	s	CE5-CBM1	3512	84	1329	70	6	3	4	X
TT 08797	Cutinase	s	CE5	2	400	6	8				
TT 09441	N-acetylglucosamine 6-phosphate deacetylase		CE9	121	35	1359	66		2	4	
TT 06499	Feruloyl esterase	s	CBM1	6276	40	10	13	9	4		X
TT 04751	Protein phosphatase 1 regulatory subunit		CBM21	278	6	218	48	3		2	
TT 00910	LysM domain-containing protein		CBM50	71	45	1	0	9	11	2	Ch
TT 06197	Endo- $\beta$ -1,4-xylanase	s	GH10	245	10	1	0	9	7	2	X
TT 08161	Endo- $\beta$ -1,4-xylanase	s	GH10-CBM1	1335	14	40	25	6	2		X
TT 09033	Endo- $\beta$ -1,4-xylanase	s	GH10	2	32	2	1	2	8	1	X
TT 01839	Endo- $\beta$ -1,4-xylanase	s	GH11	1429	8312	419	9	7	12	6	X
TT 02489	Endo- $\beta$ -1,4-xylanase	s	GH11	11	30	1	1	4	8		X
TT 03075	Endo- $\beta$ -1,4-xylanase	s	GH11-CBM1	4576	171	299	33	7	5	3	X
TT 03205	Endo- $\beta$ -1,4-xylanase	s	GH11	4	40	4	0	5	11	5	X
TT 01081	Endoglucanase	s	GH131	666	241	32	32	4	5		C
TT 09709	Feruloyl esterase	s	GH131	416	5	25	13	5	1	1	X
TT 03813	Endo-1,3(4)- $\beta$ -glucanase	s	GH16	82	4	104	15	3		3	C
TT 08907	$\beta$ -glucanase	s	GH16	32	23	139	29		2	2	C, XG
TT 04717	Chitinase	s	GH18-CBM18	0	1	13	1		2	4	Ch
TT 05685	Chitinase		GH18	8	17	189	5		4	5	Ch
TT 05010	N,O-diacetylmuramidase	s	GH25	356	249	2452	580		1	2	
TT 06031	Rhamnogalacturonase	s	GH28	201	7	6	2	7	4	2	P
TT 07104	Rhamnogalacturonase	s	GH28	0	1	0	0		5		P
TT 07602	Endopolygalacturonase	s	GH28	2	6	1	0	2	7	2	P
TT 07210	Glycosylceramidase		GH30 7	3659	78	48	7	9	6	3	
TT 02717	Trehalase		GH37	119	4	109	25	2		2	
TT 06379	Endo- $\alpha$ -1,5-L-arabinanase	s	GH43 24	120	1	56	10	4		3	XG, X, P
TT 00222	Arabinan endo- $\alpha$ -1,5-L-arabinosidase	s	GH43 30	4	0	1	0	4		2	XG, X, P
TT 02313	$\alpha$ -L-arabinofuranosidase		GH43 36	53	1	1	2	5			XG, X, P
TT 09000	Endo- $\beta$ -1,4-glucanase	s	GH45	3188	28	8	96	5			C
TT 00224	Endoglucanase	s	GH5 5	4790	27	7	7	9	4		C
TT 01019	Endo- $\beta$ -1,4-glucanase	s	GH5 5	2322	26	138	38	6	2	2	C
TT 03238	Endo- $\beta$ -1,4-glucanase	s	GH5 5	347	1	3	9	5			C
TT 05879	Endo- $\beta$ -1,4-glucanase	s	GH5 5	31	12	11	1	5	6	3	C
TT 06537	Mannan endo- $\beta$ -1,4-mannosidase	s	GH5 7	311	32	3	2	7	7		GM
TT 09640	Mannan endo- $\beta$ -1,4-mannosidase	s	GH5 7	57	6	1	0	9	8	4	GM
TT 01894	Arabinogalactan endo- $\beta$ -1,4-galactanase	s	GH53	173	77	5	6	5	6		P
TT 06655	1,4- $\beta$ -D-glucan cellobiohydrolase	s	GH6	8643	78	90	48	8	3		X
TT 09103	1,4- $\beta$ -D-glucan cellobiohydrolase	s	GH6	17	3	10	2	3	3	2	C
TT 09005	$\alpha$ -L-arabinofuranosidase	s	GH62	1	13	1	0	4	10	4	XG, X, P
TT 03032	Endoglucanase	s	GH7	170	3	7	4	6	2		C
TT 05512	Endo- $\beta$ -1,4-glucanase	s	GH7	164	2	0	0	12	8	3	C
TT 05546	Exoglucanase	s	GH7-CBM1	12001	38	26	53	8	2		C
TT 05797	Endoglucanase	s	GH7-CBM1	4654	20	2	20	8	2		C
TT 07042	1,3- $\beta$ -glucanosyltransferase	s	GH72	7	33	84	5		5	4	
TT 00146	Endo-chitosanase	s	GH75	61	1	19	2	5		3	Ch
TT 03098	Exo- $\alpha$ -L-1,5-arabinanase		GH93	94	23	122	7	4	4	4	P

log2FC 8), a CE5 acetylxylan esterase (TT\_05762 with log2FC 7) and a CE5 cutinase (TT\_08797 with log2FC 8). The two putative mannosidase encoding genes highly upregulated on Avicel (TT\_09640 with log2FC 8 and TT\_06537 with log2FC 7) were also highly upregulated on rice straw. Two putative AA7 oxidoreductase encoding genes (TT\_06681 with log2FC 4 and TT\_03025 with log2FC 5), as well as a transcript with an AA8 domain (TT\_02325 with log2FC 10), were also among the top forty highly upregulated genes on rice straw.

On beechwood xylan, upregulation of CAZyme-encoding genes was more subdued with generally lower log2FC values, and fewer overlaps with the other two substrates were detected. Interestingly, the second most highly upregulated gene encoded a putative AA9-CBM1 LPMO (TT\_01736 with log2FC 6). Transcripts of six other AA9 LPMOs were also more abundant on beechwood xylan compared to glucose (TT\_06268, log2FC 5; TT\_07962, log2FC 3; TT\_08370, log2FC 3; TT\_06407, log2FC 3; TT\_05592, log2FC 3 and TT\_09068, log2FC 3), although not as strongly upregulated as on other substrates. A gene encoding a putative AA12 PQQ-dependent pyranose dehydrogenase (TT\_09453, log2FC 5) was also among the top highly upregulated genes on BX. The same applies for a putative AA3\_2 dehydrogenase TT\_08234 with log2FC 4.

Despite beechwood xylan being a pure xylan substrate, only a fraction of upregulated CAZyme genes were encoding putative xylan-acting enzymes, such as CE5 acetylxylan esterases TT\_05762 (log2FC 6), TT\_05762 (log2FC 6) and TT\_08166 (log2FC 4), GH11 endo-1,4- $\beta$ -xylanases TT\_01839 (log2FC 6), TT\_03205 (log2FC 5) and TT\_03075 (log2FC 3). Interestingly, no GH10 endo-1,4- $\beta$ -xylanases nor CE1, CE3 or CE16 esterase encoding genes were among the most highly upregulated genes on BX. Two genes encoding enzymes that are active on chitin and possibly involved in fungal cell wall modulation were also highly upregulated on beechwood xylan, such as GH18 chitinases TT\_05685 (log2FC 5) and TT\_04717 (log2FC 4). A variety of cellulose-, pectin- and arabinan-active CAZyme-encoding genes were upregulated at low levels (log2FC 2-4); the same was observed for some genes encoding enzymes typically associated with lignin degradation (Fig. 4, Additional file 6).

## Discussion

The present study sought to explain in detail the enzymatic machinery *T. terrestris* LPH172 possesses to break down major lignocellulosic polymers based on genome and transcriptome analysis. Specifically, cellulose degradation appeared to rely mostly on LPMOs and some highly expressed canonical cellulases. Compared to other carbon sources, growth on Avicel was poor, yet LPH172

performed better on this substrate than most other fungi (Fig. 1). Poor growth on Avicel could result from lack of cellulase induction or the high crystallinity of Avicel. We noticed similar discrepancy between putative cellulose-degrading genes and poor growth on Avicel in our previous work with *M. cinnamomea* [43]. Growth discrepancies between the two *T. terrestris* strains LPH172 and CBS 117535 corroborate previously reported differences in biomass degradation and enzyme production between strains of the same species [17]. It has been reported that two strains of *A. niger*, for example, produced diverse sets of biomass-degrading enzymes, even when grown on the same plant biomass substrates [37, 60]. This was mainly attributed to the postgenomic and regulatory differences between strains. Differential gene expression analysis helped to suggest the main enzymes involved in degradation of tested substrates (Fig. 4). The range of upregulated CAZyme-encoding genes was perhaps more diverse than expected, with genes encoding mannanases, xylanases, pectinases and lignin-active enzymes being upregulated on all substrates regardless of the presence or absence of the corresponding polymers. Co-regulation of biomass-degrading enzymes or the presence of traces of other polymers could explain induction of these genes. Enzymological studies that compare the activities and activity optima of these enzymes will help determine the function of seemingly redundant enzymes such as the abundantly expressed AA9 LPMO genes.

(Hemi)cellulase encoding genes were highly expressed and upregulated on Avicel, which is a cellulosic substrate and, hence, should not require hemicellulases for degradation. However, this type of unanticipated expression has been shown before in *T. terrestris*, when the cellulosic substrate CMC induced xylanase production [16]. On the other hand, minor amounts of xylan previously reported to be found in Avicel [56, 57] could also stimulate xylanase expression. Putative direct (hemi)cellulase regulating transcription factors were analyzed in the genome of LPH172 and the presence of the prominent promiscuous regulator Xyr1 known to affect both cellulases and hemicellulases [61, 62] was detected (Additional file 4). This finding supports the hypothesis of (hemi)cellulase co-regulation in this fungus or alternatively, a combination of co-regulation and induction due to the xylan contamination in Avicel.

Since their discovery a decade ago, LPMOs have been studied in several different fungal, bacterial and even insect species, with new families and activities being continuously reported [53, 63, 64]. In *T. terrestris* LPH172, the transcriptomic data indicate that AA9 LPMOs play a crucial role in cellulose degradation, as six such enzymes were highly upregulated and five were very

highly expressed during growth on Avicel. Interestingly, on all three polymeric substrates the highest upregulated genes belonged to this family. Several AA9 genes in LPH172 were highly upregulated on rice straw, which contains some cellulose, but also on purified beechwood xylan. We hypothesize that traces of cellulose in the beechwood xylan substrate induce the expression of cellulose-degrading enzymes, or that co-regulation occurs. Alternatively, certain AA9 LPMOs could act on non-cellulosic substrates, including xylan, mannan or xyloglucan, as reported, for instance, for AA9 LPMOs from *M. cinnamomea* [65] and *M. thermophila* [66] and *N. crassa* [67]. A clear preference for CBM1-containing genes was shown among the highly expressed and upregulated CAZyme genes on Avicel but also on rice straw, supporting the predicted cellulose-binding character of CBM1 proteins.

Other members of AA CAZy families were also highly expressed and/or significantly upregulated during growth of LPH172 on various substrates. AA3\_1-AA8 cellobiose dehydrogenases (CBD) act as reducing agents to fuel LPMO reactions [23, 28, 68–71]. However, not all fungi containing LPMO genes contain supplementary cellobiose dehydrogenase encoding genes [68]. We observed high co-expression and co-upregulation of these enzyme encoding genes on cellulose-containing substrates. The AA3\_1-AA8 CBD gene (TT\_04380) that was highly co-upregulated with several AA9 LPMOs in our study has been shown to act in synergy with a *Thermoascus aurantiacus* GH61A (AA9) enzyme [28]. Interestingly, on the cellulosic substrates we also noted the upregulation of two AA8 cytochrome domain containing CAZyme genes (TT\_02325 and TT\_09190) which could also potentially reduce the copper in the active center of LPMOs. However, the electron transfer to these AA8 domains remains unclear. According to Pfam analysis, TT\_09190 also contains a putative sugar transporter domain. Moreover, absence of AA8 co-upregulation with AA9 LPMO genes on beechwood xylan might indicate that a different reduction system is utilized on hemicellulosic substrates than on the cellulosic substrates. AA3\_2 single-domain flavoenzymes have also been shown to act in synergy as electron donors for LPMOs [72]. Here, we detected co-expression and co-upregulation of an AA3\_2 dehydrogenase gene (TT\_08234) and AA9 genes on Avicel and beechwood xylan. Other AA CAZymes capable of producing  $H_2O_2$ , and therefore potentially serving as LPMO co-factors, are AA7 family oxidoreductases. In fact, an AA7 enzyme with a novel oligosaccharide dehydrogenase activity has been recently shown to both transfer electrons to the LPMO active site copper but also produce  $H_2O_2$  as a co-substrate of LPMOs [72]. Here, AA7 encoding genes TT\_06681 and TT\_03025 were upregulated on

both Avicel and rice straw; however, their exact roles in biomass degradation have to be elucidated by future studies. Interestingly, out of the 20 AA7 domains in this strain only four were expressed according to our transcriptomic analysis. An AA3 enzyme (TT\_07514) that did not fit in the top genes mentioned in our analysis in Fig. 4 but was still significantly upregulated on all substrates, is not yet classified into an AA3 sub-family according to dbCAN. This AA3 encoding gene found to contain two putative GMC-oxidoreductase domains using Pfam analysis, as well as a putative bacterial luciferase-like domain. To our knowledge, such a domain has not been seen before in combination with AA3 domains and may indicate a fifth sub-family of AA3 CAZymes, but further studies are crucial to substantiate this hypothesis.

In general, the elevated number of LPMO-encoding genes in the fungus, together with their high expression and upregulation confirm the importance of (AA9) LPMOs for plant biomass decomposition by *T. terrestris* and explains why studying the secretomes of this species had such a clear cellulase-boosting effect [19]. The numerous LPMOs in filamentous fungi support the concept of microbial mutualism. According to this concept, some fungi are responsible mainly for LPMO secretion and for attacking crystalline substrate surfaces, thereby making way for others to degrade amorphous polysaccharides and eventually benefitting the whole microbial community [60, 73]. Such interactions have been documented with regard to the mutually beneficial synthesis of vital growth substances in fungi [74]. Analogously, white rot fungi are known to degrade lignin, whereas brown rotters mainly modify lignin [7], indicating unique specifications for lignocellulose degradation in different filamentous fungi.

Finally, regarding possible genetic factors contributing to fungal thermostability [50], the genome of *T. terrestris* LPH172 revealed high GC content in the coding sequences of all genes and in those encoding CAZymes as well. In addition, the observed high GC3 content could contribute to the thermophilic lifestyle in *T. terrestris*, as also noted by Berka et al. [25]. Further research is needed to confirm and elucidate the mechanisms of this interesting phenomenon.

## Conclusion

We sequenced and analyzed the genome of a novel *T. terrestris* strain LPH172. Both genome and transcriptome analyses of the novel thermophilic *T. terrestris* strain LPH172 revealed in detail the enzymatic machinery used by the fungus to break down lignocellulosic biomass. Using transcriptome data from growth on glucose, Avicel, rice straw, and beechwood xylan we conclude that the fungus relies on an LPMO-centered strategy when grown

on cellulosic substrates and more on canonical hemicellulases when grown on xylan. The LPMO-focused degradation approach is supported by co-regulation of other AA enzyme encoding genes that likely are expressed as LPMO co-factors. We also detected high GC and GC3 content as possible genomic characteristics contributing to the thermostability of the strain. The present study provides the basis for further biochemical characterization of the lignocellulose-degrading machinery in *T. terrestris* and filamentous fungi in general. The apparent complementary or redundant nature of certain CAZymes identified in this study needs to be investigated further with enzymological techniques, whereas a more detailed physiological understanding can be achieved with additional transcriptome and proteome studies.

## Methods

### Isolation and maintenance of fungi

Samples containing decaying plant residues (compost, grasses, rice straw, mushroom ground, wood, and soil) were collected from different provinces in Northern Vietnam during 2012–2016. Fungal strains were isolated as described by Thanh et al. [17] by incubation at 50 °C and under acidic conditions (pH 2.0) on medium containing untreated rice straw as the sole carbon source. After 7–10 days of incubation, fungal colonies were transferred to potato dextrose agar (PDA) plates and purified by hyphal tip culture at 50 °C. The isolates were maintained in PDA slants in a refrigerator at 2–8 °C.

### Growth on plates

Fungal strains were streaked out on solid base medium composed of 4 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 13.6 g L<sup>-1</sup> (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.8 g L<sup>-1</sup> CaCl<sub>2</sub>·H<sub>2</sub>O, 0.6 g L<sup>-1</sup> MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.1 g L<sup>-1</sup> peptone; 0.1 g L<sup>-1</sup> yeast extract, 1000× trace element solution (10 mg L<sup>-1</sup> FeSO<sub>4</sub>·7H<sub>2</sub>O, 3.2 mg L<sup>-1</sup> MnSO<sub>4</sub>·H<sub>2</sub>O, 2.8 mg L<sup>-1</sup> ZnSO<sub>4</sub>·7H<sub>2</sub>O, 4 mg L<sup>-1</sup> CoCl<sub>2</sub>·6H<sub>2</sub>O, 3.5 mg L<sup>-1</sup> CuSO<sub>4</sub>·5H<sub>2</sub>O, pH 5.6), 1% (w/v) agar, and 2% (w/v) of one of the following carbon sources: Avicel, beechwood xylan, starch, guar gum, CMC, citrus pectin, cellobiose, D-glucose, D-xylose, locust bean gum, and inulin from Dahlia tubers or bark powder. Controls contained no carbon source. Plates were incubated at 30 °C or 50 °C for 1–7 days. Cellobiose was supplied by Megazyme. Bark powder was supplied by the Department of Chemistry and Chemical Engineering (Chalmers University of Technology, Gothenburg, Sweden) and contained 10% dried pine and 90% dried spruce bark. All other chemicals were supplied by Merck. Fungal strains were received from the collection at the Centre for Industrial Microbiology (Food Industries Research Institute, Hanoi, Vietnam).

### DNA and RNA extraction

To extract genomic DNA, strain LPH172 was grown on a PDA plate for 5 days at 50 °C, the mycelium was divided into six equal parts, and each part was used as inoculum in 100 mL liquid base medium containing 2% glucose. Cultures were incubated in 500-mL baffled Erlenmeyer flasks at 50 °C and 150 rpm for 48 h. The mycelium was harvested by filtering through sterile Miracloth (Merck Millipore) and rinsing extensively with liquid base medium without glucose. After pressing out excessive moisture by hand, the mycelium was snap-frozen in liquid nitrogen and ground to a fine powder in a Tissue-Lyser (Qiagen) at 30-s, 30-Hz intervals with pre-cooled tungsten steel balls. CTAB buffer (2% CTAB, 100 mM Tris-HCl, pH 8.0, 20 mM EDTA, 1.4 M NaCl) was immediately added at 10 mL/g<sub>mycelium</sub>, briefly vortexed and the suspension incubated at 57 °C for 1 h. DNA was purified three times by phenol–chloroform extraction until no interphase was visible, followed by 2-propanol precipitation [75]. The resulting pellet was resuspended in 1 mL TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and incubated with 200 µg mL<sup>-1</sup> RNase A (Thermo Fisher Scientific) at 60 °C for 2 h to remove residual RNA. After an additional round of phenol–chloroform extraction, the pellet was resuspended in 150 µL TE buffer and DNA was further purified with the DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. Quality of the purified DNA was verified by agarose gel electrophoresis, Nanodrop (Thermo Fisher Scientific), and Qubit Fluorometer (Thermo Fisher Scientific) before genome sequencing.

For RNA extraction, a 100-mL pre-culture on glucose was prepared as described above for DNA extraction. After harvesting and washing the mycelium, this was divided equally between 250-mL baffled Erlenmeyer flasks containing 50 mL basal liquid medium supplemented with 2% Avicel, beechwood xylan, rice straw, corn cob xylan or glucose. After 5 days of cultivation at 50 °C and 150 rpm, the mycelium was harvested, frozen, and ground to a powder, as described for DNA extraction. RNA was extracted using TRIzol (Invitrogen) and chloroform, and further purified with the RNeasy Plant RNA kit (Qiagen) with on-column DNase digestion. Quality of the purified RNA was checked by agarose gel electrophoresis, Nanodrop (Thermo Fisher Scientific), Qubit Fluorometer (Thermo Fisher Scientific) and Bioanalyzer (Agilent Technologies). High quality RNA for transcriptome sequencing had and OD 260/280 of 1.8–2.0, an OD 260/230 of 2–2.2 and an RNA integrity number (RIN) of ≥ 8. Unless otherwise mentioned, all chemicals were supplied by Merck, except for corn cob xylan (Carbosynth) and rice straw powder (Center for Industrial Microbiology).

### Genome sequencing, assembly, and analysis

Genome sequencing and assembly was carried out by GATC Biotech (Constance, Germany). According to the company's proprietary protocols, an 8–12-kb library was prepared by DNA fragmentation, size selection, end repair and adapter ligation, primer annealing, and polymerase annealing. Sequencing was performed on a PacBio RS II instrument (raw data output 400 Mb for a genome of ~ 37 Mb) with an average read length of > 6000 bp. De novo assembly of PacBio RS reads was achieved with proprietary GATC Biotech methods optimized for the sequencing technology, read length and type of raw data and included read filtering by length and quality, error correction of long PacBio reads through alignment of short reads ("reads of insert"), assembly of error corrected reads and assembly polishing. Completeness of the genome was assessed with BUSCO (v3.0.2b) against the fungi\_odb9 gene dataset (<http://buscocodev.ezlab.org/datasets/fungiodb9.tar.gz>). To analyze GC and GC3 content, seqinr, Biostrings, and sscu R packages were used [76–78].

### Transcriptome sequencing, assembly, and analysis

Transcriptome sequencing was performed by GATC Biotech (Constance, Germany) with the Inview Transcriptome Explore package. Briefly, a randomly primed cDNA library was prepared by purifying poly-A-containing mRNAs, fragmenting, adapter ligation, and PCR amplification. Illumina sequencing with single reads (50 bp) generated 24 million reads per sample that could be mapped to the reference. Quality checks were performed, and all samples achieved a percentage of clean reads > 95%. Assembly and annotation were done by National Bioinformatics Infrastructure Sweden (NBIS). Guided assembly was done with Tophat2 (v2.0.9) and Stringtie (v1.2.2), whereas repeat masking employed the Repeat-Modeler package (v1.0.8). *Ab initio* training for annotation was done with GeneMark-ET (v4.3), Augustus, and snap. Gene builds were computed using the MAKER package (v3.01.1), which employed the following software: exonerate (v2.4), Blast+ (v2.2.28), RepeatMasker (v4.0.3), Bioperl (v1.6.922), Augustus (v2.7), tRNAscan (v1.3.1), snap, and GeneMark-ET (v4.3). Functional annotation of genes and transcripts was performed using the translated CDS features of each coding transcript. For each predicted protein sequence, a BLASTp search was performed on the UniProt/Swiss-Prot reference dataset with default parameters (e-value cut-off = 1, similarity cut-off = 30%) to retrieve gene name and protein function. Secreted proteins were predicted using the SignalP 4.0 Server. Genes containing CAZy domains were identified using dbCAN2 (accessed October 2019). Reads

were assigned and counted to the genome annotation using featureCounts of the Rsubread package (v2.2.6) in R and converted to TPMs. Differential gene expression was analyzed using edgeR (v3.30.3) [59] in R with TMM-normalization and removal of reads with less than 1 read per million in all samples.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13068-021-01975-1>.

**Additional file 1:** Homologous sequences of transcription-elongation factor-alpha and beta-tubulin genes used for identifying the strain as *T. terrestris* LPH172 and phylogenetic analysis for the strain identification.

**Additional file 2:** Growth of *T. terrestris* LPH172 and other biomass-degrading filamentous fungi on different carbon sources.

**Additional file 3:** The complete list of putative CAZy domains detected in *T. terrestris* LPH172 and other filamentous fungi.

**Additional file 4:** Putative homologues of known transcription factors directly regulating (hemi)cellulose utilization in the genome of *T. terrestris* LPH172.

**Additional file 5:** The complete list of all expressed genes on the four tested substrates.

**Additional file 6:** The complete list of all upregulated putative CAZyme encoding genes on the four tested substrates.

### Acknowledgements

Support from National Bioinformatics Infrastructure Sweden (NBIS) is gratefully acknowledged. We thank Professor Jens Christian Frisvad (Technical University of Denmark) for his help with identifying the strain. We also thank Martin Engqvist and Sinisa Bratulic (Chalmers University of Technology) for their help with the script used for GC(3) content calculations.

### Authors' contributions

SH, LO, VNT conceptualized the study. SH and NTT performed the experiments under supervision of LO, VNT and JL. MT, PR and SH performed the data analyses. MT and SH wrote the manuscript. All authors contributed to results discussions, read and approved the final manuscript. All authors have read and approved the final manuscript.

### Funding

Open access funding provided by Chalmers University of Technology. The collaboration between FIRI and Chalmers was funded by The Swedish Research Council Dnr 348-2014-3523. LO, SH, MT and JL acknowledge the financial support to them from Knut and Alice Wallenberg Foundation and Chalmers Foundation for their activities within Wallenberg Wood Science Center. PR acknowledges the funding from the Novo Nordisk Foundation grant NNF19OC0055044.

### Availability of data and materials

All data generated or analyzed during this study are included within the article and as Additional files 1, 2, 3, 4, 5 and 6. The genome assembly has been deposited at DDBJ/EMBL/GenBank under the assembly Accession No. GCA\_900343105.1. Transcriptomic data are deposited at the European Nucleotide Archive (ENA) under the Accession No. PRJEB25201 (ERP107096). The strain is deposited in the Food Industries Research Institute (Hanoi, Vietnam) culture collection.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Wallenberg Wood Science Centre, Department of Biology and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. <sup>2</sup>Division of Industrial Biotechnology, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden. <sup>3</sup>Center for Industrial Microbiology, Food Industries Research Institute, Thanh Xuan, Hanoi, Vietnam.

Received: 14 October 2020 Accepted: 18 May 2021

Published online: 03 June 2021

**References**

- Hassan SS, Williams GA, Jaiswal AK. Lignocellulosic biorefineries in Europe: Current state and prospects. *Trends Biotechnol*. 2019;37(3):231–4.
- Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady J, et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*. 2007;315(5813):804–7.
- Van Dyk JS, Pletschke BI. A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes-factors affecting enzymes, conversion and synergy. *Biotechnol Adv*. 2012;30(6):1458–80.
- Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons BA, Blanch HW. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol Bioeng*. 2012;109(4):1083–7.
- Adrio JL, Demain AL. Microbial enzymes: tools for biotechnological processes. *Biomolecules*. 2014;4(1):117–39.
- Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, et al. Fueling the future with fungal genomics. *Mycology*. 2011;2(3):192–209.
- Mäkelä MR, Donofrio N, De Vries RP. Plant biomass degradation by fungi. *Fungal Genet Biol*. 2014;72:2–9.
- Patel AK, Singhanian RR, Sim SJ, Pandey A. Thermostable cellulases: current status and perspectives. *Bioresour Technol*. 2019;279:385–92.
- Atalah J, Cáceres-Moreno P, Espina G, Blamey JM. Thermophiles and the applications of their enzymes as new biocatalysts. *Bioresour Technol*. 2019;280:478–88.
- Viikari L, Alapuranen M, Puranen T, Vehmaanperä J, Siika-aho M. Thermostable enzymes in lignocellulose hydrolysis. In: Olsson L, editor. *Biofuels*. Berlin: Springer; 2007. p. 121–45.
- Krška D, Larsbrink J. Investigation of a thermostable multi-domain xylanase-glucuronoyl esterase enzyme from *Caldicellulosiruptor kristjanssonii* incorporating multiple carbohydrate-binding modules. *Biotechnol Biofuels*. 2020;13(1):1–13.
- Wang XW, Bai FY, Bensch K, Meijer M, Sun BD, Han YF, et al. Phylogenetic re-evaluation of *Thielavia* with the introduction of a new family *Podosporaceae*. *Stud Mycol*. 2019;93:155–252.
- Margaritis A, Merchant RFJ. Thermostable cellulases from thermophilic microorganisms. *Crit Rev Biotechnol*. 1986;4(3):327–67.
- Durand H. Comparative study of cellulases and hemicellulases from four fungi: mesophiles *Trichoderma reesei* and *Penicillium* sp. and thermophiles *Thielavia terrestris* and *Sporotrichum cellophilum*. *Enzyme Microb Technol*. 1984;6:175–80.
- Yang S, Xu H, Yan Q, Liu Y, Zhou P, Jiang Z. A low molecular mass cutinase of *Thielavia terrestris* efficiently hydrolyzes poly(esters). *J Ind Microbiol Biotechnol*. 2013;40(2):217–26.
- García-Huante Y, Cayetano-Cruz M, Santiago-Hernández A, Cano-Ramírez C, Marsch-Moreno R, Campos JE, et al. The thermophilic biomass-degrading fungus *Thielavia terrestris* Co3Bag1 produces a hyperthermophilic and thermostable  $\beta$ -1,4-xylanase with exo- and endo-activity. *Extremophiles*. 2017;21(1):175–86.
- Thanh VN, Thuy NT, Huong HTT, Hien DD, Hang DTM, Anh DTK, et al. Surveying of acid-tolerant thermophilic lignocellulolytic fungi in Vietnam reveals surprisingly high genetic diversity. *Sci Rep*. 2019;9(1):1–12.
- Benoit JB, Yoder JA, Zettler LW, Hobbs HH. Mycoflora of a troglodytic cave cricket, *Hadenocerus cumberlandicus* (Orthoptera: *Rhaphidophoridae*), from two small caves in Northeastern Kentucky. *Ann Entomol Soc Am*. 2004;97(5):989–93.
- Merino ST, Cherry J. Progress and challenges in enzyme development for biomass utilization. In: Olsson L, editor. *Biofuels*. Berlin: Springer; 2007. p. 95–120.
- Harris PV, Welner D, McFarland KC, Re E, Navarro Poulsen JC, Brown K, et al. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: Structure and function of a large, enigmatic family. *Biochemistry*. 2010;49(15):3305–16.
- Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen JCN, Johansen KS, et al. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc Natl Acad Sci USA*. 2011;108(37):15079–84.
- Westereng B, Ishida T, Vaaje-Kolstad G, Wu M, Eijsink VGH, Igarashi K, et al. The putative endoglucanase PcGH61D from *Phanerochaete chrysosporium* is a metal-dependent oxidative enzyme that cleaves cellulose. *PLoS ONE*. 2011. <https://doi.org/10.1371/journal.pone.0027807>.
- Phillips CM, Beeson WT, Cate JH, Marletta MA. Cellobiose dehydrogenase and a copper-dependent polysaccharide monooxygenase potentiate cellulose degradation by *Neurospora crassa*. *ACS Chem Biol*. 2011;6(12):1399–406.
- Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels*. 2013;6(1):1–14.
- Berka RM, Grigoriev IV, Otillar R, Salamov A, Grimwood J, Reid I, et al. Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat Biotechnol*. 2011;29(10):922–7.
- de Vries RP, Benoit I, Doehlemann G, Kobayashi T, Magnuson JK, Panisko EA, et al. Post-genomic approaches to understanding interactions between fungi and their environment. *IMA Fungus*. 2011;2(1):81–6.
- Gilbert M, Yaguchi M, Watson DC, Wong KKY, Breuil C, Saddler JN. A comparison of two xylanases from the thermophilic fungi *Thielavia terrestris* and *Thermoascus crustaceus*. *Appl Microbiol Biotechnol*. 1993;40(4):508–14.
- Langston JA, Brown K, Xu F, Borch K, Garner A, Sweeney MD. Cloning, expression, and characterization of a cellobiose dehydrogenase from *Thielavia terrestris* induced under cellulose growth conditions. *Biochim Biophys Acta*. 2012;1824(6):802–12.
- Xu H, Yan Q, Duan X, Yang S, Jiang Z. Characterization of an acidic cold-adapted cutinase from *Thielavia terrestris* and its application in flavor ester synthesis. *Food Chem*. 2015;188:439–45.
- Woon JS, Mackeen MM, Binsudin AH, Mahadi NM, Illias RM, Murad AM, et al. Production of an oligosaccharide-specific cellobiohydrolase from the thermophilic fungus *Thielavia terrestris*. *Biotechnol Lett*. 2016;38(5):825–32.
- Meng Z, Yang QZ, Wang Zhen J, Hou YH. Cloning, characterization, and functional expression of a thermostable type B feruloyl esterase from thermophilic *Thielavia terrestris*. *Appl Biochem Biotechnol*. 2019;189(4):1304–17.
- Rodríguez-Mendoza J, Santiago-Hernández A, Alvarez-Zúñiga MT, Gutiérrez-Antón M, Aguilar-Orsorio G, Hidalgo-Lara ME. Purification and biochemical characterization of a novel thermophilic exo- $\beta$ -1,3-glucanase from the thermophile biomass-degrading fungus *Thielavia terrestris* Co3Bag1. *Electron J Biotechnol*. 2019;41:60–71.
- Shirke AN, Basore D, Holton S, Su A, Baugh E, Butterfoss GL, et al. Influence of surface charge, binding site residues and glycosylation on *Thielavia terrestris* cutinase biochemical characteristics. *Appl Microbiol Biotechnol*. 2016;100(10):4435–46.
- Banerjee S, Archana A, Satyanarayana T. Xylanolytic activity and xylose utilization by thermophilic molds. *Folia Microbiol*. 1995;40(3):279–82.
- Andersen MR, Salazar MP, Schaap PJ, van de Vondervoort PJ, Culley D, Thykaer J. Comparative genomics of citric-acid-producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88. *Genome Res*. 2011;21(6):885–97.
- Xue M, Yang J, Li Z, Hu S, Yao N, Dean RA, et al. Comparative analysis of the genomes of two field isolates of the rice blast fungus *Magnaporthe oryzae*. *PLoS Genet*. 2012. <https://doi.org/10.1371/journal.pgen.1002869>.

37. de Vries RP, Riley R, Wiebenga A, Aguilar-Osorio G, Amillis S, Uchima CA, et al. Comparative genomics reveals high biological diversity and specific adaptations in the industrially and medically important fungal genus *Aspergillus*. *Genome Biol.* 2017;18:1–45.
38. Money NP. Fungal diversity. In: Watkinson SC, Boddy L, Money NP, editors. *The Fungi*. Cambridge: Academic Press; 2016. p. 1–36.
39. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–2.
40. Mohanta TK, Bae H. The diversity of fungal genome. *Biol Proced Online.* 2015;17(1):1–9.
41. Stajich JE. Fungal genomes and insights into the evolution of the kingdom. *Microbiol Spectr.* 2017;5(4):619–33.
42. Kjørboelling I, Vesth T, Frisvad JC, Nybo JL, Theobald S, Kildgaard S, et al. A comparative genomics study of 23 *Aspergillus* species from section Flavi. *Nat Commun.* 2020. <https://doi.org/10.1038/s41467-019-14051-y>.
43. Hüttner S, Nguyen TT, Granchi Z, Chin-A-Woeng T, Ahrén D, Larsbrink J, et al. Combined genome and transcriptome sequencing to investigate the plant cell wall degrading enzyme system in the thermophilic fungus *Malbranchea cinnamomea*. *Biotechnol Biofuels.* 2017;10(1):265.
44. Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, et al. The genome sequence of the model ascomycete fungus *Podospira anserina*. *Genome Biol.* 2008. <https://doi.org/10.1186/gb-2008-9-5-r77>.
45. Floudas D, Binder M, Riley R, Barry K, Blanchette R, Henrissat B, et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science.* 2012;336:1715–9.
46. Ohm RA, De Jong JF, Lugones LG, Aerts A, Kothe E, Stajich JE, et al. Genome sequence of the model mushroom *Schizophyllum commune*. *Nat Biotechnol.* 2010;28(9):957–63.
47. Hüttner S, Granchi Z, Nguyen TT, van Pelt S, Larsbrink J, Thanh VN, et al. Genome sequence of Rhizomucor pusillus FCH 57, a thermophilic zygomycete involved in plant biomass degradation harbouring putative GH9 endoglucanases. *Biotechnol Reports.* 2018;20:e00279.
48. Ma LJ, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, et al. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet.* 2009. <https://doi.org/10.1371/journal.pgen.1000549>.
49. Van Noort V, Bradatsch B, Arumugam M, Amlacher S, Bange G, Creevey C, et al. Consistent mutational paths predict eukaryotic thermostability. *BMC Evol Biol.* 2013. <https://doi.org/10.1186/1471-2148-13-7>.
50. Salar RK. Origin of thermophily in fungi. In: Salar RK, editor. *Thermophilic Fungi*. Boca Raton: CRC Press; 2018. p. 29–53.
51. Chang YC, Tsai HF, Karos M, Kwon-Chung KJ. THTA, a thermotolerance gene of *Aspergillus fumigatus*. *Fungal Genet Biol.* 2004;41(9):888–96.
52. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014;42(D1):490–5.
53. Filiatrault-Chastel C, Navarro D, Haon M, Grisel S, Herpoël-Gimbert I, Chevreton D, et al. AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. *Biotechnol Biofuels.* 2019;12(1):55.
54. Benocci T, Victoria M, Pontes A, Zhou M, Seiboth B, de Vries RP. Regulators of plant biomass degradation in ascomycetous fungi. *Biotechnol Biofuels.* 2017;10(152):1–25.
55. Yokoyama S, Matsura Y. The Asian Biomass Handbook. The Japan Institute of Energy. (2008). p 338
56. van Gool MP. Targeted discovery and functional characterisation of complex-xylan degrading enzymes. [Wageningen]: Wageningen University; 2012.
57. Várnai A, Siika-aho M, Viikari L. Restriction of the enzymatic hydrolysis of steam-pretreated spruce by lignin and hemicellulose. *Enzyme Microb Technol.* 2010;46(3–4):185–93.
58. Mazurkewich S, Poulsen JCN, Lo LL, Larsbrink J. Structural and biochemical studies of the glucuronoyl esterase OtCE15A illuminate its interaction with lignocellulosic components. *J Biol Chem.* 2019;294(52):19978–87.
59. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2009;26(1):139–40.
60. de Vries RP, Mäkelä MR. Genomic and postgenomic diversity of fungal plant biomass degradation approaches. *Trends Microbiol.* 2020;28(6):487–99.
61. Klaubauf S, Narang HM, Post H, Zhou M, Brunner K, Mach-Aigner AR, et al. Similar is not the same: differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi. *Fungal Genet Biol.* 2014;72:73–81.
62. Stricker AR, Grosstessner-Hain K, Würleiner E, Mach RL. Xyr1 (Xylanase Regulator 1) regulates both the hydrolytic enzyme system and D-xyllose metabolism in *Hypocrea jecorina*. *Eukaryot Cell.* 2006;5(12):2128–37.
63. Sabbadin F, Hemsworth GR, Ciano L, Henrissat B, Dupree P, Tryfona T, et al. An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nat Commun.* 2018. <https://doi.org/10.1038/s41467-018-03142-x>.
64. Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, et al. Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. *Nat Chem Biol.* 2018;14(3):306–10.
65. Hüttner S, Várnai A, Petrovic D, Bach CX, Kim Anh DT, Thanh VN, et al. Specific Xylan Activity Revealed for AA9 Lytic Polysaccharide Monooxygenases of the Thermophilic Fungus *Malbranchea cinnamomea* by Functional Characterization. *Appl Environ Microbiol.* 2019;85(23):1–13.
66. Frommhagen M, Sforza S, Westphal AH, Visser J, Hinz SWA, Koetsier MJ, et al. Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. *Biotechnol Biofuels.* 2015;8(1):4–15.
67. Petrović DM, Várnai A, Dimarogona M, Mathiesen G, Sandgren M, Westergren B, et al. Comparison of three seemingly similar lytic polysaccharide monooxygenases from *Neurospora crassa* suggests different roles in plant biomass degradation. *J Biol Chem.* 2019;294(41):15068–81.
68. Kracher D, Scheiblbrandner S, Felice AKG, Breslmayr E, Preims M, Ludwig K, et al. Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science.* 2016;352(6289):1098–101.
69. Tan TC, Kracher D, Gandini R, Sygmund C, Kittl R, Haltrich D, et al. Structural basis for cellobiose dehydrogenase action during oxidative cellulose degradation. *Nat Commun.* 2015;6(1):7542.
70. Courtade G, Wimmer R, Røhr ÅK, Preims M, Felice AKG, Dimarogona M, et al. Interactions of a fungal lytic polysaccharide monooxygenase with β-glucan substrates and cellobiose dehydrogenase. *Proc Natl Acad Sci USA.* 2016;113(21):5922–7.
71. Loose JSM, Forsberg Z, Kracher D, Scheiblbrandner S, Ludwig R, Eijssink VGH, et al. Activation of bacterial lytic polysaccharide monooxygenases with cellobiose dehydrogenase. *Protein Sci.* 2016;25(12):2175–86.
72. Haddad Momeni M, Fredslund F, Bissaro B, Raji O, Vuong TV, Meier S, et al. Discovery of fungal oligosaccharide-oxidising flavo-enzymes with previously unknown substrates, redox-activity profiles and interplay with LPMOs. *Nat Commun.* 2021;12(1):2132.
73. Boddy L. Interactions between fungi and other microbes. In: Watkinson SC, Boddy L, Money NP, editors. *The Fungi*. Boston: Academic press; 2016.
74. Kögl F, Fries N. Über den Einfluß von Biotin, Aneurin und Meso-Inositol auf das Wachstum verschiedener Pilzarten. 26. Mitteilung über pflanzliche Wachstumsstoffe. *Hoppe Seylers Z Physiol Chem.* 1937;249(2–4):93–110.
75. Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol:chloroform. *Cold Spring Harb Protoc.* 2006. <https://doi.org/10.1101/pdb.prot4455>.
76. Charif D, Lobry JR. SeqinR 1.0-2: A contributed package to the R Project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. *Structural Approaches to Sequence Evolution*. Berlin: Springer; 2007. p. 207–32.
77. Pagès H, Aboyou P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. 2020
78. Sun Y. sscu: Strength of Selected Codon Usage. 2020

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.