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

Monika Tõlgo
Chalmers University of Technology: Chalmers tekniska hogskola
Silvia Hüttner
Chalmers University of Technology https://orcid.org/0000-0002-7096-9680
Nguyen Than Thuy
Food Industries Research Institute
Vu Nguyen Than
Food Industries Research Institute
Johan Larsbrink
Chalmers University of Technology: Chalmers tekniska hogskola
Lisbeth Olsson (lisbeth.olsson@chalmers.se)
Chalmers University of Technology: Chalmers tekniska hogskola https://orcid.org/0000-0002-0827-5442

Research

Keywords: biomass degradation, carbohydrate active enzymes, cellulose, filamentous fungi, LPMO, thermostable enzymes, transcriptome, xylan.

Posted Date: October 26th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-96298/v1

Version of Record: A version of this preprint was published at Biotechnology for Biofuels on June 3rd, 2021. See the published version at https://doi.org/10.1186/s13068-021-01975-1.

Abstract

Background: Biomass-degrading enzymes with improved activity and stability can ameliorate 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 well 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 analysed the 36.6-Mb genome and transcriptome of LPH172 during growth on glucose, cellulose, rice straw, and beechwood xylan. In total, 411 CAZy domains were found among 10,128 predicted genes. Compared to other fungi, auxiliary activity (AA) enzymes were particularly enriched. GC content was higher in coding sequences than in the overall genome. A high GC3 content was hypothesised to contribute to thermophilicity. *T. terrestris* employed mainly lytic polysaccharide monooxygenases (LPMOs) and glycoside hydrolase (GH) family 7 glucanases to attack cellulosic substrates, and conventional hemicellulases (GH10 and GH11) to degrade xylan. The observed co-expression and co-upregulation of AA9 LPMOs, other AA CAZymes, and (hemi)cellulases points to a complex and nuanced degradation strategy. Growth on more complex and heterogeneous substrates resulted in a more varied but generally lower gene expression.

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

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 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 elevated 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 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 *T. 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. Thermophilic enzymes can lower industrial processing costs as they can achieve faster reaction rates, greater stability, and are more easily adjustable to various set-ups (10).

Thielavia terrestris (syn Thermothielavioides terrestris (11)*)* is a well-known filamentous fungus identified in 1983 as a potential source of thermostable industrial enzymes based on successful (hemi)cellulase assays (12, 13). The species is a thermophilic saprobic Ascomycete isolated mainly from soil and compost in Asia (14–16) and from a cave cricket species in North America (17). As described by Merino and Cherry (18), *T. terrestris* played a pivotal role in the discovery of lytic polysaccharide monooxygenases (LPMOs). Cultivation broth from *T. terrestris* primed for cellulase production showed striking synergy in degrading pretreated corn stover when supplemented with the enzyme cocktail Celluclast. This experiment eventually contributed to the discovery of glycoside hydrolase family 61 (GH61) proteins (18–22), today known as auxiliary activity family 9 (AA9) (23). 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 genome sequences (24). 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 (25) and detailed biochemical characterisation (14, 15, 19, 26–32), *T. terrestris* produces an array of biomass-degrading enzymes. However, no study has elucidated the gene expression mechanisms underlying the fungus' lignocellulolytic machinery or its (hemi)cellulase regulatory system 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 (16, 33–36). Here, we set out to sequence and analyse the genome and transcriptome of the newly isolated *T. terrestris* strain LPH172, which is characterised by superior enzymatic activity, thermostability, and pH stability (16). Our current study aimed to elucidate the fungus' lignocellulose degrading machinery 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 likely relied mainly on LPMOs when grown on cellulosic substrates; whereas on hemicellulosic substrates, more conventional hemicellulases were induced. Interestingly, we also report co-expression and co-upregulation between LPMOs and other AA enzymes.

Results

Strain identification

We previously isolated *T. terrestris* strain LPH172 from compost in Northern Vietnam and showed that it could be exploited as an industrially relevant enzyme producer (16). To confirm the identity of the fungus,

here, we used two common fungal household genes (37) encoding transcription elongation factor 1-*α* and *β*-tubulin. The homologous gene sequences used for the identification procedure are listed in Additional File 1. LPH172 transcription elongation factor 1-*α* was 99.93% identical (one nucleotide difference) to its *T. terrestris* NRRL 8126 homologue, while *β*-tubulin showed 100% identity. 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 utilise 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 *β*-glucosidases. Direct comparison to the previously sequenced *T. terrestris* CBS 117535 (38) showed that LPH172 grew slightly better on most substrates except glucose. Good growth was observed on pectin and inulin (a fructose polymer); whereas growth on locust bean gum and guar gum (galactomannans), as well as bark powder was poor (Additional File 2).

Genome characterisation

General features

The genome of *T. terrestris* LPH172 was sequenced on a PacBio RS II instrument by GATC Biotech (Konstanz, Germany); it yielded 527 523 reads comprising over 7 billion bases. Table 1 gives an overview of the *T. terrestris* LPH172 genome. Its size was determined to be 36.6 Mb and guanine and cytosine (GC) content was 52.60%. The latter differs from the 54.80% value reported in GenBank, possibly due to differences in calculations, such as inclusion of “N” nucleotides. 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.

Table 1
Overview of *T. terrestris* LPH172 genome.

Genome assembly	
Number of nucleotides	36,579,697
GC content	52.60%
N50 (bp)	3,006,457
Number of protein coding genes	10,128
Average gene length (bp)	1,628
Average CDS length (bp)	1,355
Average number of exons per gene	3
Average exon length (bp)	460
Longest genes (bp)	25,297
Longest exons (bp)	9,745
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	8,879
Genes without functional annotation	1,249
Genes annotated (BLASTp, e-value < $1e^{-6}$)	6,114

The size of fungal genomes can vary by orders of magnitude, the average for Ascomycota is 36.91 Mb (40, 41). Table 2 gives a brief overview of LPH172 genome characteristics compared to other industrial and lignocellulose-degrading fungi with varied origin and thermostability. Even though the genome of *T. terrestris* NRRL 8126 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. LPH172 genome size 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 1,628 bp and the average coding sequence was 1,355 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 < $1e^{-6}$).

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

Organism	Phylum	Temperature preference	Genome (Mb)	Genome GC (%)	Number of protein-coding genes	Source
Aspergillus oryzae RIB40	Ascomycota	Mesophilic	37.9	47.2	12,074	(42)
Myceliophthora thermophila ATCC 42464		Thermophilic	38.7	51.4	9,110	(24)
Malbranchea cinnamomea FCH 10.5		Thermophilic	25.0	49.8	9,437	(43)
Thielavia terrestris LPH172		Thermophilic	36.6	52.6	10,128	this article
Thielavia terrestis NRRL 8126		Thermophilic	36.9	54.7	9,813	(24)
Gloeophyllum trabeum ATCC 11539	Basidiomycota	Mesophilic	37.2	52.9	11,755	(44)
Podospora anserina S mat+		Mesophilic	35.0	52.0	10,545	(45)
Schizophyllum commune H4-8		Mesophilic	38.5	N/A	13,210*	(46)
Rhizomucor pusillus FCH 5.7	Zygomycota	Thermophilic	25.6	45.0	10,898	(47)
Rhizopus oryzae 99–880		Mesophilic	39.1	35.4	17,467	(48)
*nr of genes						

Thermostability features

Although there is no clear consensus on the causes contributing to elevated optimum temperature 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) (24, 50).

In contrast with the thermophilic ascomycete *Malbranchea cinnamomea* (43), the genome of *T. terrestris* LPH172 was not smaller compared to that of other mesophilic fungi (Table 2). GC content in gene-coding sequences was 62.0%, which was higher than the genome average of 52.6%. When looking only at the subset of genes encoding CAZymes, the average GC content was even higher (64.5%). 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 $1.3e^{-19}$) to the known thermotolerance gene *THTA* from *Aspergillus fumigatus* (GenBank: AY560012.1) (51).

CAZyme comparison with other fungi

Plant biomass-degrading and other CAZymes are catalogued into families and subfamilies in the Carbohydrate Active enZymes (CAZy) database (http://www.cazy.org/) (52).The number of CAZy domains and distribution across different CAZy families in *T. terrestris* LPH172 was analysed and compared to other known fungal biomass degraders to assess the propensity for lignocellulose degradation (Table 3). Note that by 'CAZymes' in this article we mean individual CAZyme domains. In total, 411 individual CAZy domains were detected in LPH172 using dbCAN2 (HMMER algorithm). Most CAZy domains were found to be GHs (201 candidates), with GH16 (n = 14), GH18 (n = 15), GH3 (n = 12), and GH47 (n = 10) being the most abundant subfamilies. There were also 86 glycosyl transferases (GTs), 4 polysaccharide lyases (PLs), 26 carbohydrate esterases (CEs), 83 AAs, and 11 carbohydrate-binding modules (CBMs). Compared to strain NRRL 8126, two more GHs (one GH16 and one GH47) were identified in LPH172, as well as one additional AA12, one GT2, and one CE1 (Additional File 3). *T. terrestris* LPH172 had a relatively low number of PLs compared to other fungi (Fig. 2), but a larger complement of AA family enzymes, particularly AA9 (n = 18), AA8 (n = 3), and AA7 (n = 20) (Fig. 3). Five members of AA11 (chitin-cleaving) LPMOs were detected in both *T. terrestris* strains, but no AA13 (starch-cleaving LPMOs) or AA14 (xylan-cleaving LPMOs) members were observed. LPH172 and NRRL 8126 were the only fungi, among the ones selected, presenting an AA16, a recently characterised C1-hydroxylating LPMO (53). The number of multidomain CAZymes was low; only 15 LPH172 proteins had two predicted CAZy domains, and one had three (Additional File 3).

Putative candidates for CAZymes capable of degrading all major lignocellulosic polymers (cellulose, xylan, xyloglucan, (galacto)glucomannan, pectin, and lignin), as well as starch, inulin, and chitin were found. This finding was in line with growth of *T. terrestris* on all of these carbon sources (Additional File 2).

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

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

GH, glycoside hydrolase; GT, glycoside transferase; AA, auxiliary activity; CE, carbohydrate esterase; PL, polysaccharide lyase; CBM, carbohydrate-binding module. All CAZy domains were identified using dbCAN2 (HMMER algorithm).

Regulation of plant cell wall-degrading enzymes

Regulation of (hemi)cellulolytic enzymes in filamentous fungi occurs mainly at the transcriptional level (54–56). Here, we used BLASTn and BLASTp to detect possible homologues of known transcription factors (TFs) from regulatory cascades recorded in other filamentous fungi. TF genes related to lignocellulose degradation in *T. terrestris* LPH172 included transcriptional (hemi)cellulase activator XYR1/XLNR1 (TT_07823), cellulase activators Clr-1 (TT_06796) and Clr-2 (TT_06838), known carbon-catabolite repressor CreA (TT_07794), cellulase repressor ACE1 (TT_01416), and arabinose-responsive Ara1 (TT_09773). A homology search revealed the presence of positive cellulase regulator McmA (TT_02138), C-derepressing VIB1 (TT_03515), and Hap-complex protein Hap5 (TT_04392).

Transcriptome analysis

Highly expressed genes on Avicel, rice straw, and beechwood xylan

To verify genome annotation and analyse gene expression, in particular CAZyme-encoding gene expression, the transcriptome was analysed 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 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 (57) 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 (58, 59), was selected to identify enzymes required to degrade a highly crystalline and recalcitrant cellulosic substrate. Transcriptome data from RNAseq experiments were 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.

Table 4
Forty most highly expressed genes during *T. terrestris* LPH172 growth on Avicel.

Transcript ID	fpkm	CAZy domain(s)	Putative function
TT_06621	6586	-	NA
TT_06050	6458	-	NA
TT_00578	4005	-	Respiratory supercomplex factor 2 homolog
TT_05797	3343	GH7-CBM1	Endoglucanase
TT_03518	2876	-	NA
TT_08370	2417	AA9	Endo-*β*-1,4-glucanase
TT_06655	2353	GH6	1,4-*β*-D-glucan cellobiohydrolase
TT_03075	2251	GH11-CBM1	Endo-1,4-*β*-xylanase
TT_06499	2110	CBM1	Feruloyl esterase
TT_09215	2012	-	Lactose permease
TT_00215	1665	-	Oxidoreductase
TT_07455	1622	AA9	LPMO
TT_07008	1467	-	NA
TT_08166	1450	CE5-CBM1	Acetylxylan esterase
TT_05599	1337	-	Mitochondrial oxidase
TT_00225	1326	AA4	Vanillyl-alcohol oxidase
TT_10132	1309	-	Cytochrome c
TT_09465	1232	-	Cross-pathway control protein 1
TT_09870	1089	-	Protein FDD123
TT_05357	1073	-	Acyl-CoA desaturase
TT_06750	1049	-	NA
TT_00529	1019	-	NA
TT_07123	940	-	NA
TT_01736	931	AA9	LPMO
TT_04350	928	AA9-CBM1	LPMO
TT_03837	880	-	5'-AMP-activated protein kinase subunit
TT_05536	815	-	Elongation factor 3

Transcript ID	fpkm	CAZy domain(s)	Putative function
TT_09000	790	GH45	Endoglucanase
TT_04380	777	AA3_1-AA8	Cellobiose dehydrogenase
TT_06689	755	-	Inositol oxygenase
TT_00703	682	-	SDO1-like protein
TT_01019	679	GH5_5	Endoglucanase
TT_10041	665	-	Actin-related protein
TT_00207	640	-	Voltage-gated potassium channel subunit
TT_03870	630	-	Multiprotein-bridging factor
TT_09312	613	-	Protein vip1
TT_05898	611	-	NA
TT_06609	609	-	Uncharacterized protein C32A11.02c
TT_07036	571	-	Transcriptional regulatory protein
TT_08478	557	-	Histone H2B

Fpkm values indicate average normalized transcript levels from three replicates. CAZy domains were predicted by dbCAN2 and functions were annotated by BLASTp search against the UniProt/Swiss-Prot reference dataset. NA, not annotated.

Table 5
Forty most highly expressed genes during *T. terrestris* LPH172 growth on rice straw.

Transcript ID	fpkm	CAZy domain(s)	Putative function
TT_10132	10224	-	Cytochrome c
TT_06693	8469	-	Stress protein DDR48
TT_06050	7056	-	NA
TT_06689	4893	-	Inositol oxygenase
TT_08478	4739	-	Histone H2B
TT_02247	3189	-	Mitochrondrial valine–tRNA ligase
TT_00469	2782	-	60S ribosomal protein
TT_09215	2701	-	Lactose permease
TT_01345	2474	-	40S ribosomal protein
TT_02932	2465	-	60S ribosomal protein
TT_01839	2461	GH11	Endo-1,4-*β*-xylanase
TT_01967	2044	-	60S ribosomal protein
TT_01009	1893	-	NA
TT_04612	1849	-	40S ribosomal protein
TT_00107	1730	-	NA
TT_02482	1678	-	NA
TT_02213	1642	-	Elongation factor 1-*α*
TT_01072	1633	-	60S ribosomal protein
TT_07670	1577	-	Peptide chain release factor 1
TT_00703	1534	-	SDO1-like protein C21C3.19
TT_02715	1465	-	NA
TT_02172	1448	-	Translation initiation factor
TT_06668	1436	-	Hedgehog-interacting protein
TT_00918	1410	-	Mitochondrial eptidyl-prolyl cis-trans isomerase
TT_02974	1404	-	Heat shock protein
TT_08110	1381	-	Glycogen phosphorylase
TT_01225	1358	-	THO complex subunit 4A

Transcript ID	fpkm	CAZy domain(s)	Putative function
TT_09947	1341	-	Mitochondrial phosphate carrier protein
TT_02608	1327	-	DNA-binding protein
TT_02868	1321	-	60S ribosomal protein
TT_01583	1318	-	40S ribosomal protein
TT_05389	1233	-	Ubiquitin-60S ribosomal protein
TT_04052	1181	-	60S ribosomal protein
TT_00966	1179	-	Allergen Asp f 4
TT_03265	1149	-	Tropomyosin
TT_06621	1129	-	NA
TT_01563	1093	-	Polypeptide-associated complex subunit α
TT_00802	1077	-	40S ribosomal protein
TT_08653	1059	-	Translation initiation factor 3 subunit C
TT_01895	1046	-	40S ribosomal protein

Fpkm values indicate average normalized transcript levels from three replicates. CAZy domains were predicted by dbCAN2 and functions were annotated by BLASTp search against the UniProt/Swiss-Prot reference dataset. NA, not annotated.

Table 6
Forty most highly expressed genes during *T. terrestris* LPH172 growth on beechwood xylan.

Transcript ID	fpkm	CAZy domain(s)	Putative function
TT_05599	2440	-	Mitochondrial oxidase
TT_00578	2053	-	Respiratory supercomplex factor 2
TT_03518	2024	-	NA
TT_10132	1967	-	Cytochrome c
TT_05357	1748	-	Acyl-CoA desaturase
TT_06621	1234	-	NA
TT_02482	1054	-	NA
TT_05010	1054	GH25	N,O-diacetylmuramidase
TT_01009	1030	-	NA
TT_05536	1028	-	Elongation factor 3
TT_07036	1009	-	Transcriptional regulatory protein
TT_07008	1007	-	NA
TT_06689	1004	-	Inositol oxygenase
TT_09947	913	-	Mitochondrial phosphate carrier protein
TT_09076	894	-	Copper-containing nitrite reductase
TT_00107	869	-	NA
TT_09465	867	-	Cross-pathway control protein 1
TT_03837	822	-	5'-AMP-activated protein kinase subunit β-2
TT_09870	802	-	Protein FDD123
TT_06824	796	-	Heat shock 70 kDa protein
TT_03870	795	-	Multiprotein-bridging factor
TT_09441	754	CE9	N-acetylglucosamine-6-phosphate deacetylase
TT_04420	753	-	Uncharacterized protein C18H10.17c
TT_08478	748	-	Histone H2B
TT_02247	710	-	Mitochondrial valine–tRNA ligase
TT_06668	690	-	Hedgehog-interacting protein

Transcript ID	fpkm	CAZy domain(s)	Putative function
TT_09930	665	-	Histone H3
TT_06609	634	-	Uncharacterized protein C32A11.02c
TT_04469	632	-	5-methyltetrahydropteroyltriglutamate–homocysteine methyltransferase
TT_00469	590	-	60S ribosomal protein
TT_02932	576	-	60S ribosomal protein
TT_04772	573	-	Melanoma-associated antigen
TT_01967	562	-	60S ribosomal protein
TT_06693	558	-	Stress protein DDR48
TT_08166	554	CE5-CBM1	Acetylxylan esterase
TT_03035	552	GH72	1,3-*β*-glucanosyltransferase
TT_01345	535	-	40S ribosomal protein
TT_08034	534	-	AN1-type zinc finger protein
TT_01037	534	-	Glycerol-3-phosphate dehydrogenase
TT_01225	525	-	THO complex subunit 4A

Fpkm values indicate average normalized transcript levels from three replicates. CAZy domains were predicted by dbCAN2 and functions were annotated by BLASTp search against the UniProt/Swiss-Prot reference dataset. NA, not annotated.

To identify which genes, including CAZyme-encoding genes, were the most highly expressed (by transcript number) on the chosen substrates, we looked at the top 40 (arbitrary number) candidates under each growth condition, ranked by their average fragments per kilobase million (fpkm) value. The complete list of all expressed genes is available in Additional File 4. In general, the fpkm values of the 40 most abundant transcripts varied from 6586 to 557 for growth on Avicel (Table 4), from 10,224 to 1,046 for growth on rice straw (Table 5), and from 2,440 to 525 for growth on beechwood xylan (Table 6). Interestingly, when grown on Avicel, two of the most highly expressed genes encoded short peptides of 22 (TT_06621) and 124 (TT_06050) amino acids. TT_06621 was also among the top 40 expressed genes on both rice straw and beechwood xylan, whereas TT_06050 was very highly expressed on rice straw but not on beechwood xylan or glucose. The fourth most highly expressed gene on Avicel encoded a CAZyme: a putative GH7 endoglucanase with a CBM1 (TT_05797). Twelve other putative CAZymes were among the 40 most abundant transcripts on Avicel. These included typical cellulose-active enzymes, such as four AA9 LPMOs (TT_08370, TT_07455, TT_01736, and TT_04350), a GH6 cellobiohydrolase (TT_06655),

GH5 and GH45 endoglucanases (TT_09000 and TT_01019), and an AA3-AA8 cellobiose dehydrogenase (TT_04380), as well as typical xylan-active enzymes, such as a GH11-CBM1 endoxylanase (TT_03075), a BLAST-annotated feruloyl esterase with a CBM1 (TT_06499), and a CE5 acetyl-xylan esterase (TT_08166). CAZymes with cellulose-binding CBM1 domains were overrepresented among the 40 most highly expressed genes on Avicel, and included all five CAZymes with a CBM1 in *T. terrestris* LPH172. Interestingly, a lactose permease (TT_09215) was also very abundant on both Avicel and rice straw.

Growth on rice straw seemed to favour gene expression and translation processes, as indicated by the high number of ribosome- and histone-related gene products. Out of 40 highly expressed genes, 12 encoded ribosomal subunits, which could coincide with the generally higher fpkm values on rice straw. Given the diverse polymer composition of this substrate, it was surprising to see only one CAZyme among the top 40 transcripts – a GH11 endo-1,4-*β*-xylanase (TT_01839) (Table 5). AA9 LPMOs, xylanases, and acetyl xylanesterases were also expressed on this substrate, but at lower levels (Additional File 4). A stress-response (TT_06693) and heat shock protein homologue (TT_02974) were highly expressed on rice straw, suggesting stress conditions during growth.

On beechwood xylan, four CAZymes were found among the 40 most highly expressed genes: a GH25 N,*O*-diacetyl muramidase (TT_05010), a CE9 N-acetylglucosamine-9-phosphate deacetylase (TT_09441), a CE5-CBM1 acetylxylan esterase (TT_08166), and a GH72 1,3-*β*-glucanosyltransferase. Three of those CAZymes are not involved in lignocellulosic biomass degradation but in growth and remodelling of the fungal cell wall (GH72) (60), fungal amino sugar metabolism during chitin degradation (CE9) (61), and defence against bacteria (GH25) (62). Similar to growth on rice straw, most highly expressed genes on beechwood xylan were related to general cellular metabolism (e.g. mitochondrial proteins), gene expression (histones and ribosomal proteins), and stress response (heat shock proteins and stress proteins). Biomass-degrading CAZymes, such as xylanases (GH10 and GH11), mannosidases (GH76), AA2 and AA3 oxidoreductases, and a GH13 amylase, were generally expressed at lower levels on beechwood xylan (Additional File 4).

Upregulated CAZymes on Avicel, rice straw, and beechwood xylan

Gene expression levels do not show the full spectrum of available lignocellulose-degrading enzymes in the organism, because many of them are sufficiently active at low concentration. Therefore, to understand which genes were induced under the tested conditions (Avicel, rice straw or beechwood xylan), we examined the differential expression of CAZymes with respect to glucose as reference. In particular, we focused on transcripts that were significantly more abundant (i.e. upregulated) compared to glucose.

On Avicel, AA9 LPMOs in combination with a AA3-AA8 cellobiose dehydrogenase and GH7, GH5, and GH45 endoglucanases, were the most highly upregulated CAZymes (Fig. 4, Additional File 5). AA9 LPMOs showed the highest differential expression, particularly in the case of TT_01736 (4644-fold), TT_08370 (1793-fold), and TT_04350 (468-fold). All three enzymes presented also very high fpkm numbers, indicating both high upregulation and high expression levels. Four more AA9 LPMOs were also

significantly more abundant on Avicel: TT_07455 (84-fold), TT_06268 (64-fold), TT_04352 (43-fold), and TT_02354 (5-fold). Notably, TT_06268 exhibited an fpkm value of only 5, whereas the other AA9s had fpkm values between 25 and 2417. Interestingly, many non-cellulose-acting CAZymes were also upregulated on Avicel; this included feruloyl esterase (TT_06499, 525-fold), GH11 xylanases (TT_03075, 151-fold; TT_01839, 102-fold; TT_08161, 30-fold), CE16 and CE5 acetyl (xylan) esterases (TT_06012, 147-fold; TT_08166, 54-fold; TT_05762, 26-fold), and an AA4 vanillyl-alcohol oxidase (TT_00225, 426-fold). The expression levels of these genes varied widely.

The set of highly upregulated CAZymes on rice straw shared some candidates with Avicel, such as several AA9 LPMOs (TT_06268, 1134-fold; TT_08370, 265-fold; TT_04352, 150-fold; TT_01736, 140-fold) and the hemicellulose-active GH11 xylanases TT_01839 (1951-fold) and TT_02489 (105-fold), CE5 acetylxylan esterase TT_05762 (55-fold), and CE16 acetyl esterase TT_06012 (52-fold). Generally, more hemicellulose-active enzymes were highly upregulated on rice straw than on Avicel, pointing to a more diverse substrate composition of the former. Interestingly, the highest upregulation on rice straw was detected for the mannan endo-1,4-*β*-mannosidase TT_06537 (1951-fold), even though mannan is not a major polymer in this substrate. Notably, this gene had a low fpkm value of 53. The second most upregulated gene, GH11 endo-1,4-*β*-xylanase TT_01839 (1610-fold), had an fpkm of 2461. A few putative cellulose-acting enzymes were upregulated on rice straw but not on Avicel, such as AA9 LPMO TT_03770 (12-fold) and the AA8 TT_09190 (19-fold). Another putative AA8 cellobiose dehydrogenase (TT_02325) was upregulated 1168-fold, although not at a statistically significant level ($p = 0.135$) (Fig. 4, Additional File 5).

On beechwood xylan, upregulation of CAZymes was more muted, and fewer overlaps with other substrates were detected. Despite beechwood xylan being a pure xylan substrate, only a fraction of upregulated CAZymes were xylan-acting, such as CE5 acetylxylanesterases TT_05762 (67-fold) and TT_08166 (21-fold), GH11 endo-1,4-*β*-xylanases TT_01839 (35-fold) and TT_03075 (10-fold), and GH10 endo-1,4-*β*-xylanases TT_08161 (5-fold) and TT_09033 (5-fold). Acetylxylan esterase TT_05762 presented the highest upregulation and expression on beechwood xylan; whereas the other candidates were more highly upregulated and expressed on Avicel, rice straw, or both. Several enzymes active on chitin and possibly involved in fungal cell wall modulation were upregulated on beechwood xylan, such as GH18 chitinases TT_05685 (28-fold), TT_04717 (11-fold), endo-chitosanoase TT_08109 (3-fold), and the GH72 and CE9 enzymes mentioned above. Transcripts of several AA9 LPMOs were also more abundant on beechwood xylan compared to glucose (TT_06268, 36-fold; TT_01736, 36-fold; TT_08370, 5-fold), although again at much lower levels than on the other substrates. A variety of cellulose-, mannan-, pectin- and arabinan-active CAZymes were upregulated at low levels (2- to 4-fold); the same was observed for some enzymes typically associated with lignin degradation (Fig. 4, Additional File 5).

Discussion

The present study sought to explain in detail the enzymatic machinery *T. terrestris* LPH172 possessed to break down major lignocellulosic polymers based on genome and transcriptome analysis. Specifically,

cellulose degradation relied mostly on LPMOs and some highly expressed endoglucanases. 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 elevated crystallinity of Avicel. 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 (16). Recently, de Vries and Mäkelä reported that related fungi with similar genomic content produced highly diverse sets of enzymes, even when grown on the same plant biomass substrates. Future studies should examine whether such differences originate from instability of fungal genomes (63).

Our analysis highlighted the presence of a homologue of promiscuous (hemi)cellulolytic regulator XYR1/XLNR1/XlnR in the genome of *T. terrestris*. Clr-1 and Clr-2 are known as essential cellulolytic TFs in *Neurospora crassa* (64). While the genes are conserved among filamentous Ascomycetes, their functionality is only partly conserved as reviewed by Benocci et al. (56). Therefore, TFs in *T. terrestris* could co-regulate both cellulases and hemicellulases. Alternatively, cross-talk between regulatory pathways could ensure that the reaction products of some cellulases are responsible also for hemicellulase induction. (Hemi)cellulase co-regulation is supported by abundant xylanase expression and upregulation on Avicel, which is a cellulosic substrate and, hence, does not require hemicellulases for degradation. This type of unanticipated regulation was shown before in *T. terrestris*, when the cellulosic substrate CMC was used to induce xylanase production (15), although minor amounts of xylan in Avicel (58, 59) could also stimulate xylanases.

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 (53, 65, 66) reported. In *T. terrestris* LPH172, AA9 LPMOs play a crucial role in cellulose degradation, as six such enzymes were highly upregulated and four were very highly expressed during growth on Avicel. What remains to be determined is why only some of the overall 18 AA9s are upregulated on this substrate, whereas six showed no and five only very low expression. It is possible that certain AA9 enzymes are induced by or active only on a subset of cellulose, or on entirely different substrates (67, 68). Several AA9s in LPH172 were highly upregulated on rice straw, which contains some cellulose, but also on beechwood xylan, which is made purely of xylan. We hypothesize that traces of cellulose in the substrate induce the expression of cellulose-degrading enzymes, or that co-regulation occurs. Alternatively, certain AA9 LPMOs could act on non-celluose substrates, including xylan, mannan or xyloglucan, as in the thermophilic fungus *M. cinnamomea* (67). A clear preference for CBM1-containing genes was shown among the upregulated CAZymes on Avicel, supporting the cellulose-binding character of CBM1.

Interestingly, the majority of CAZymes necessary for hemicellulose degradation, such as most xylanases, were only expressed at comparatively low levels (Tables 5 and 6). It is possible that certain CAZymes do not require high expression levels to be sufficiently active; whereas others, such as several AA9 LPMOs, need to be induced in higher amounts. Why CAZyme expression was generally so much higher on Avicel compared to rice straw or beechwood xylan remains to be determined. A stronger induction response on

cellulosic substrate is likely not the reason, as rice straw also contains cellulose. Elevated expression of stress- and translation-related genes on rice straw and beechwood xylan points to possibly higher demands on the cell compared to growth on Avicel. However, the idea that these substrates are less suitable for growth is refuted by the results on solid media, which showed better growth on beechwood xylan than on Avicel (Fig. 1). One explanation relates to the higher stress encountered by cells during growth in shake flasks, as performed here for transcriptomics analysis. Because under these conditions mixing of nutrients, pH, and oxygen saturation are not controlled, cells may activate a stress response. To discard this possibility, experiments should be conducted in a highly controlled environment such as bioreactors. Absence of CAZyme expression on rice straw in particular could arise from use of non-pretreated substrate, which was insufficiently accessible to biomass-degrading enzymes and for inducer molecules to activate the CAZyme machinery.

On the one hand, the highly crystalline Avicel may require elevated amounts of a few CAZymes to be broken down, compared to the more accessible and less crystalline polymers in rice straw and beechwood xylan. On the other hand, a complex substrate like rice straw needs a more active translation machinery (i.e. ribosomes) to produce a wider response to the different polymers present, as indicated here by elevated expression of ribosomal proteins.

Differential gene expression analysis helped identify the main enzymes involved in degradation of tested substrates (Fig. 4). The range of upregulated CAZymes was perhaps more diverse than expected, with mannanases, xylanases, 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 contamination with traces of other polymers could explain induction of these genes. Similar to AA9 LPMOs, not all members in a CAZy family were upregulated or expressed to the same degree; however, the same major variation was observed for GH10 and GH11 xylanases. Incomplete gene duplications during fungal evolution could result in truncated genes that still contain a CAZy domain but are not transcribed. A more fine-tuned regulation of gene expression depending on substrates and conditions is also likely. Comparison of gene expression under a wider range of conditions will elucidate the above possibilities. Enzymological studies that compare the activities and activity optima of these enzymes will help determine the function of seemingly redundant enzymes. Initially, the transcriptome analysis in the present study included corn cob xylan as an arabinoxylan-containing substrate model for cereals. The results were not included, because the purchased corn cob xylan turned out to be composed only of xylo-oligomers.

AA3_1-AA8 cellobiose dehydrogenases work as reducing agents to fuel LPMO reactions (22, 27, 69–72). Here, we observed high co-expression and co-upregulation of these enzymes on cellulose-containing substrates. The AA3_1-AA8 CBD (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) (27). Moreover, absence of such co-expression on beechwood xylan might indicate how LPMOs were probably involved only in the degradation of cellulosic substrates rather than xylan. AA3_2 single-domain flavoenzymes have also been shown to act in synergy as electron donors for LPMOs (73). Here, we

detected co-expression and co-upregulation of AA3_2 s and AA9s. Three AA3_2 s (TT_08234, TT_05138, and TT_05809) were upregulated on Avicel, one on rice straw (TT_5138), and one on beechwood xylan (TT_8234). Transcripts of these AA3s were not very abundant, possibly indicating sufficient activity even at low concentrations.

Co-expression, co-upregulation or synergy between AA9s and AA4 vanillyl-alcohol oxidase has not been reported before to the best of our knowledge and has been noticed for the first time in this study. The AA4 TT_00225 was highly expressed and upregulated on Avicel and upregulated on rice straw but not on beechwood xylan. This enzyme catalyses the oxidation of vanillyl alcohol to vanillin with the release of hydrogen peroxide (74). Vanillyl alcohol could result also from lignin degradation, which may explain upregulation of this enzyme on rice straw. Its upregulation on Avicel could, instead, be explained by co-regulation of cellulolytic and lignocellulolytic enzymes. Either way, AA4s can produce H_2O_2, which is also a co-substrate for LPMOs (75–77). In the case of high LPMO expression and catalysis, it is conceivable that the fungus tries to produce enough co-substrate for all its LPMOs. However, we could find neither a match for this AA4 in a *T. terrestris* secretome (24), nor a putative signal peptide for TT_00225. Other AA CAZymes capable of producing H_2O_2, and therefore potentially serving as LPMO co-factors, are AA7 glucooligosaccharide oxidases. Here, the AA7 TT_6681 was upregulated on both Avicel and rice straw; however, its role in biomass degradation will be detailed by future studies.

An AA3 enzyme (TT_07514), not yet classified into an AA3 sub-family according to dbCAN, was found to contain two putative GMC-oxireductase domains with 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. Luciferases are classified as oxidoreductases, and a homologue of luciferase-like monooxygenase has been shown to be the most abundant protein in *Escherichia coli* when grown on vanillin (78). In our study, TT_07514 was also upregulated on rice straw, which contains lignin, supporting the participation of luciferase-like domains in oxidative cleavage of lignin and/or (hemi)celluloses.

In general, the elevated number of LPMO-encoding genes in the fungus, together with their high expression and upregulation, confirm the importance of LPMOs for decomposition by *T. terrestris*. Compared to the transcriptome of *M. cinnamomea* (43) where one AA11 was upregulated when grown on wheat bran compared to glucose and another AA11 upregulated when comparing growth on wheat bran with growth on xylan. No AA11 LPMOs were highly expressed or upregulated in LPH172, even though its genome contains five AA11s. Another poorly expressed gene in *T. terrestris* encoded an AA16 LPMO (53). The numerous LPMOs in filamentous fungi support the concept of microbial mutualism. Accordingly, some fungi could be responsible mainly for LPMO secretion and attacking crystalline substrate surfaces, making way for others to degrade amorphous polysaccharides and eventually benefitting the whole fungal community (63, 79). Such interactions have been documented with regard to the mutually beneficial synthesis of vital growth substances in fungi (80). Analogously, white rot fungi are known to degrade lignin, whereas brown rotters are only capable of modifying lignin (7), indicating unique specifications for lignocellulose degradation in different filamentous fungi.

Another interesting protein found upregulated in our study was the lactose permease TT_09215. We hypothesise it might be linked to the regulation of plant cell wall degradation as lactose has been shown to induce (hemi)cellulase production in *T. reesei* (81, 82). It is possible that the permease is promiscuous, and it is used for transporting other (di)saccharides which would be reasonable in respect of our study. The 124-aa-long peptide TT_6050 was highly expressed both on Avicel and rice straw, but not on beechwood xylan or glucose. The same peptide was also detected in the alfalfa and barley straw secretome of *T. terrestris* NRRL 8126 (24), confirming its importance for lignocellulose degradation by *T. terrestris.*

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 particularly in those encoding for CAZymes. Additionally, the observed high GC3 content could contribute to the thermophilic lifestyle in *T. terrestris* as also noted by Berka et al. (24).

Conclusion

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-centred strategy when grown on cellulosic substrates. This approach is supported by co-regulation of other AA enzymes that likely serve as LPMO co-factors. We also observed that more crystalline substrates required a different CAZyme expression strategy than the heterogeneous rice straw and the less recalcitrant beechwood xylan. The present study provides the basis for further biochemical characterisation 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. (16) 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^{-1} KH_2PO_4, 13.6 g L^{-1} $(NH_4)_2SO_4$, 0.8 g L^{-1} $CaCl_2·H_2O$, 0.6 g L^{-1} $MgSO_4·7H_2O$, 0.1 g L^{-1} peptone; 0.1 g L^{-1} yeast extract, 1000 × trace element solution (10 mg L^{-1} $FeSO_4·7H_2O$, 3.2 mg L^{-1} $MnSO_4·H_2O$, 2.8 mg L^{-1} $ZnSO_4·7H_2O$, 4 mg L^{-1} $CoCl_2·6H_2O$, 3.5 mg L^{-1} $CuSO_4·5H_2O$, 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 Center 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 TissueLyser (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 $mL/g_{mycelium}$, 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 (83). 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^{-1} 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 (Life Technologies) 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 RNAeasy Plant RNA kit (Qiagen) with on-column DNAse digestion. Quality of the purified RNA was checked by agarose gel electrophoresis, Nanodrop, and Qubit Fluorometer. 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 was carried out by GATC Biotech. 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 methods and included read filtering by length and quality, error correction, alignment of short reads, and assembly polishing. Completeness of the genome was assessed with BUSCO (v3.0.2b) against the fungi_odb9 gene dataset. To analyse GC and GC3 content, seqinr, Biostrings, and sscu R packages were used (84–86). To determine the presence of homologous genes such as TFs, BLASTn and BLASTp were used to align candidates with LPH172 coding sequences or protein sequences, respectively. For BLASTn, minimum query coverage of 19% and minimum sequence similarity of 60% were used as indicators of homology. For further confirmation, BLASTp cut-off values were set to minimum 50% query coverage and 45% sequence similarity. Only significant hits were analysed ($p \leq 0.05$).

Transcriptome sequencing, assembly, and analysis

Transcriptome sequencing was performed by GATC Biotech 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 30 million reads. Quality checks, 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 RepeatModeler 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. Secretory proteins were predicted using the SignalP 4.0 Server. Genes containing CAZy domains were identified using dbCAN2 (accessed February 2018).

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.

Funding

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.

Authors' contributions

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

Acknowledgements

Support from National Bioinformatics Infrastructure Sweden (NBIS) is gratefully acknowledged. We also thank Professor Jens Christian Frisvad from Technical University of Denmark for his help with identifying the strain.

References

1. Hassan SS, Williams GA, Jaiswal AK. Lignocellulosic Biorefineries in Europe: Current State and Prospects. Trends Biotechnol. 2019;37(3):231–4.
2. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady J., et al. Biomass Recalcitrance: Engineering Plants and Enzymes for Biofuels Production. Science (80-). 2007;315(5813):804–7.
3. 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.
4. 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.
5. Adrio JL, Demain AL. Microbial enzymes: tools for biotechnological processes. Biomolecules. 2014;4(1):117–39.
6. Grigoriev I V, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, et al. Fueling the future with fungal genomics. Mycology. 2011;2(3):192–209.
7. Mäkelä MR, Donofrio N, De Vries RP. Plant biomass degradation by fungi. Fungal Genet Biol. 2014;72:2–9.
8. Patel AK, Singhania RR, Sim SJ, Pandey A. Thermostable cellulases: Current status and perspectives. Bioresour Technol. 2019;279(November 2018):385–92.

9. Atalah J, Cáceres-Moreno P, Espina G, Blamey JM. Thermophiles and the applications of their enzymes as new biocatalysts. Bioresour Technol. 2019;280(February):478–88.
10. Viikari L, Alapuranen M, Puranen T, Vehmaanperä J, Siika-aho M. Thermostable Enzymes in Lignocellulose Hydrolysis. In: Olsson L, editor. Biofuels. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 121–45.
11. 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(August):155–252.
12. Margaritis A, Merchant RFJ. Thermostable Cellulases from Thermophilic Microorganisms. Crit Rev Biotechnol. 1986;4(3):327–67.
13. Durand H. Comparative study of cellulases and hemicellulases from four fungi : mesophiles Trichoderma ree i and Penicillimn sp . and thermophiles Thiela via terrestris and Sporotdch um ceilnlophilmn. 1984;6:175–80.
14. 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.
15. 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 β-1,4-xylanase with exo- and endo-activity. Extremophiles. 2017;21(1):175–86.
16. 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.
17. Benoit JB, Yoder JA, Zettler LW, Hobbs HH. Mycoflora of a trogloxenic cave cricket, Hadenoecus cumberlandicus (Orthoptera: Rhaphidophoridae), from two small caves in Northeastern Kentucky. Ann Entomol Soc Am. 2004;97(5):989–93.
18. Merino ST, Cherry J. Progress and Challenges in Enzyme Development for Biomass Utilization. In: Olsson L, editor. Biofuels. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 95–120.
19. Harris P V., 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.
20. 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 U S A. 2011;108(37):15079–84.
21. 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;6(11).
22. 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.

23. 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.
24. Berka, Randy M and Grigoriev, Igor V and Otillar, Robert and Salamov, Asaf and Grimwood, Jane and Reid, Ian and Ishmael, Nadeeza and John, Tricia and Darmond, Corinne and Moisan M-C and others. Comparative genomic analysis of the thermophilic biomass-degrading fungi Myceliophthora thermophila and Thielavia terrestris. Nat Biotechnol. 2011;29(10):922–7.
25. 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.
26. 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.
27. 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 - Proteins Proteomics. 2012;1824(6):802–12.
28. 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.
29. Woon JSK, Mackeen MM, Sudin AH bin, Mahadi NM, Illias RM, Murad AMA, et al. Production of an oligosaccharide-specific cellobiohydrolase from the thermophilic fungus Thielavia terrestris. Biotechnol Lett. 2016;38(5):825–32.
30. Meng Z, Yang QZ, Wang J zhen, 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.
31. Rodríguez-Mendoza J, Santiago-Hernández A, Alvarez-Zúñiga MT, Gutiérrez-Antón M, Aguilar-Osorio G, Hidalgo-Lara ME. A novel thermophilic exo-β-1,3-glucanase from the thermophile biomass-degrading fungus Thielavia terrestris Co3Bag1. Purification and biochemical characterization of the enzyme. Electron J Biotechnol. 2019;41:60–71.
32. 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.
33. Banerjee S, Archana A, Satyanarayana T. Xylanolytic activity and xylose utilization by thermophilic molds. Folia Microbiol (Praha). 1995;40(3):279–82.
34. Andersen MR, Salazar MP, Schaap PJ. Comparative genomics of citric-acid-producing Aspergillus niger ATCC 1015 versus enzyme-producing CBS 513.88. Genome Res. 2011;21(6):885–97.
35. 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;8(8).
36. 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. Vol. 18, Genome Biology. 2017. 1–45 p.

37. Money NP. Chapter 1 - Fungal Diversity. In: Watkinson SC, Boddy L, Money NP, editors. The Fungi. Third. Boston: Academic Press; 2016. p. 1–36.
38. Powell AJ. Thielavia terrestris CBS 117535 Resequencing (Project ID: 1004727) [Internet]. 2012. Available from: https://genome.jgi.doe.gov/portal/Thitersequencing_3_FD/Thitersequencing_3_FD.info.html
39. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 Oct;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. The Fungal Kingdom. 2017;5(4):619–33.
42. Machida M, Asai K, Sano M, Tanaka T, Kumagai T, Terai G, et al. Genome sequencing and analysis of Aspergillus oryzae. Nature. 2005;438(7071):1157–61.
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. 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 (80-). 2012;336(June):1715–9.
45. Espagne E, Lespinet O, Malagnac F, Da Silva C, Jaillon O, Porcel BM, et al. The genome sequence of the model ascomycete fungus Podospora anserina. Genome Biol. 2008;9(5).
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 5.7, 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;5(7).
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;13(1).
50. Salar RK. Origin of Thermophily in Fungi. In: Thermophilic Fungi. 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, Chevret D, et al. AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. Biotechnol Biofuels. 2019;12(1):55.
54. Aro N, Pakula T, Penttilä M. Transcriptional regulation of plant cell wall degradation by filamentous fungi. FEMS Microbiol Rev. 2005;29(4):719–39.
55. Kunitake E, Kobayashi T. Conservation and diversity of the regulators of cellulolytic enzyme genes in Ascomycete fungi. Curr Genet. 2017;63(6):951–8.
56. Benocci T, Victoria M, Pontes A, Zhou M, Seiboth B, Vries RP De. Biotechnology for Biofuels Regulators of plant biomass degradation in ascomycetous fungi. Biotechnol Biofuels. 2017;1–25.
57. Yokoyama S, Matsmura Y. The Asian Biomass Handbook. 2008. 338 p.
58. van Gool MP. Targeted discovery and functional characterisation of complex-xylan degrading enzymes. 2012.
59. 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.
60. Samalova M, Mélida H, Vilaplana F, Bulone V, Soanes DM, Talbot NJ, et al. The β-1,3-glucanosyltransferases (Gels) affect the structure of the rice blast fungal cell wall during appressorium-mediated plant infection. Cell Microbiol. 2017;19(3):1–14.
61. Vincent F, Yates D, Garman E, Davies GJ, Brannigan JA. The three-dimensional structure of the N-acetylglucosamine-6-phosphate deacetylase, NagA, from Bacillus subtilis: A member of the urease superfamily. J Biol Chem. 2004;279(4):2809–16.
62. Korczynska JE, Danielsen S, Schagerlöf U, Turkenburg JP, Davies GJ, Wilson KS, et al. The structure of a family GH25 lysozyme from Aspergillus fumigatus. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2010;66(9):973–7.
63. de Vries RP, Mäkelä MR. Genomic and Postgenomic Diversity of Fungal Plant Biomass Degradation Approaches. Vol. 28, Trends in Microbiology. Elsevier Ltd; 2020. p. 487–99.
64. Craig JP, Coradetti ST, Starr TL, Glass NL. Direct Target Network of the Neurospora crassa Plant Cell Wall. MBio. 2015;6(5):1–11.
65. 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;9(1).
66. 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.
67. Hüttner S, Várnai A, Petrovic DM. 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.
68. Petrović DM, Várnai A, Dimarogona M, Mathiesen G, Sandgren M, Westereng 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.

69. Kracher D, Scheiblbrandner S, Felice AKG, Eijsink VGH, Ludwicka K, Felice AKG, et al. Extracellular electron transfer systems fuel cellulose oxidative degradation. Science (80-). 2016;352(6289):1098–101.
70. 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(May).
71. 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. 2016;113(21):5922–7.
72. Loose JSM, Forsberg Z, Kracher D, Scheiblbrandner S, Ludwig R, Eijsink VGH, et al. Activation of bacterial lytic polysaccharide monooxygenases with cellobiose dehydrogenase. Protein Sci. 2016;25(12):2175–86.
73. Garajova S, Mathieu Y, Beccia MR, Bennati-Granier C, Biaso F, Fanuel M, et al. Single-domain flavoenzymes trigger lytic polysaccharide monooxygenases for oxidative degradation of cellulose. Sci Rep. 2016;6(June):1–9.
74. de Jong E, Berkel WJH van B, Zwan RP van der Z, Bont JAM de B. Purification and characterization of vanillyl-alcohol oxidase from. Eur J Biochem. 1992;657:651–7.
75. Müller G, Chylenski P, Bissaro B, Eijsink VGH, Horn SJ. The impact of hydrogen peroxide supply on LPMO activity and overall saccharification efficiency of a commercial cellulase cocktail. Biotechnol Biofuels [Internet]. 2018;11(1):1–17. Available from: https://doi.org/10.1186/s13068-018-1199-4
76. Bissaro B, Røhr ÅK, Müller G, Chylenski P, Skaugen M, Forsberg Z, et al. Oxidative cleavage of polysaccharides by monocopper enzymes depends on H2O2. Nat Chem Biol. 2017;13(10):1123–8.
77. Jones SM, Transue WJ, Meier KK, Kelemen B, Solomon EI. Kinetic analysis of amino acid radicals formed in H2O2-driven CuI LPMO reoxidation implicates dominant homolytic reactivity. Proc Natl Acad Sci U S A. 2020;117(22):2–8.
78. Pattrick C, Webb JP, Green J, Chaudhuri RR, Collins MO, Kelly J. Vanillin Toxicity in Escherichia coli. 2019;4(4):1–29.
79. Boddy L. Chapter 10 - Interactions Between Fungi and Other Microbes. In: Watkinson SC, Boddy L, Money NP, editors. The Fungi. Third. Boston: Academic Press; 2016. p. 337–60.
80. Kögl F, Fries N. Über den Einfluß von Biotin, Aneurin und Meso-Inosit auf das Wachstum verschiedener Pilzarten. 26. Mitteilung über pflanzliche Wachsturnsstoffe. Hoppe Seylers Z Physiol Chem. 1937;249(2–4):93–110.
81. Penttilä M. Heterologous protein production in Trichoderma. In: Trichoderma and gliocladium 2. 1998. p. 365–82.
82. Zhang F, Bunterngsook B, Li J-X, Zhao X-Q, Verawat C, Chen-Guang L, et al. Regulation and production of lignocellulolytic enzymes from Trichoderma reesei for biofuels production. Adv Bioenergy. 2019;4:79–119.

83. Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol:chloroform. Cold Spring Harb Protoc. 2006;
84. 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: Molecules, Networks, Populations. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 207–32.
85. Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. 2020.
86. Sun Y. sscu: Strength of Selected Codon Usage. 2020.

Figures

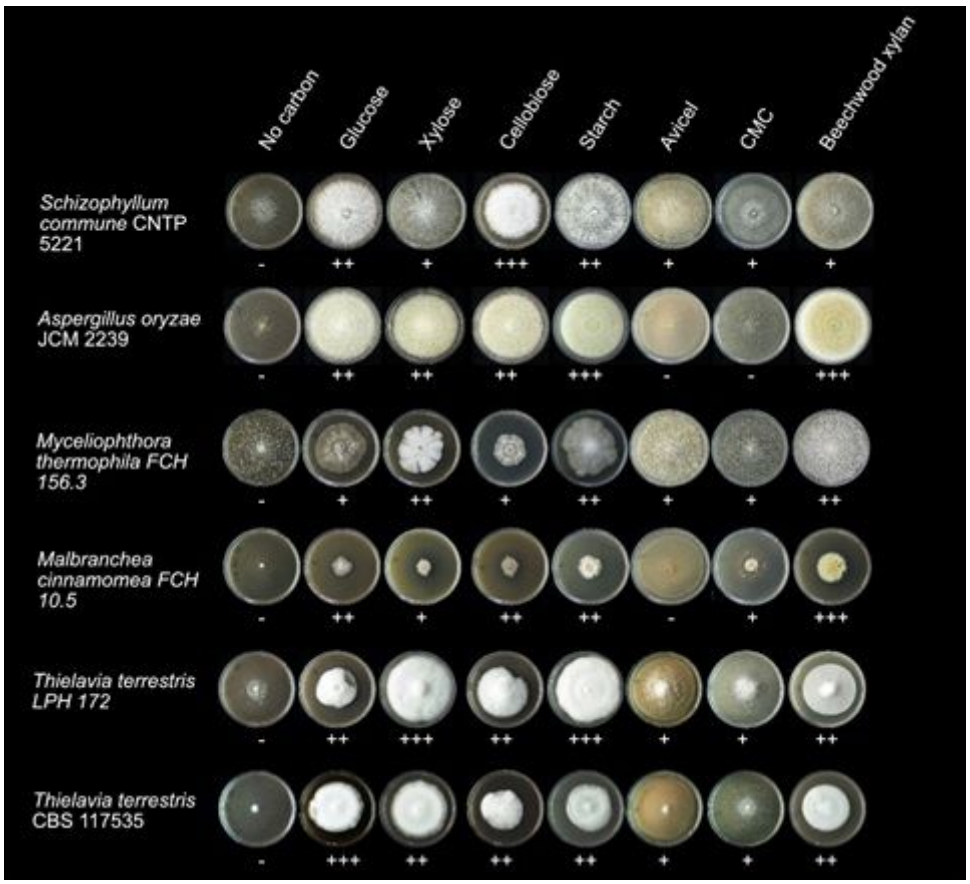

Figure 1

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 7 days. Growth was evaluated visually and categorized from - (no growth) to +++ (very good growth), depending on the extent and density of the mycelium.

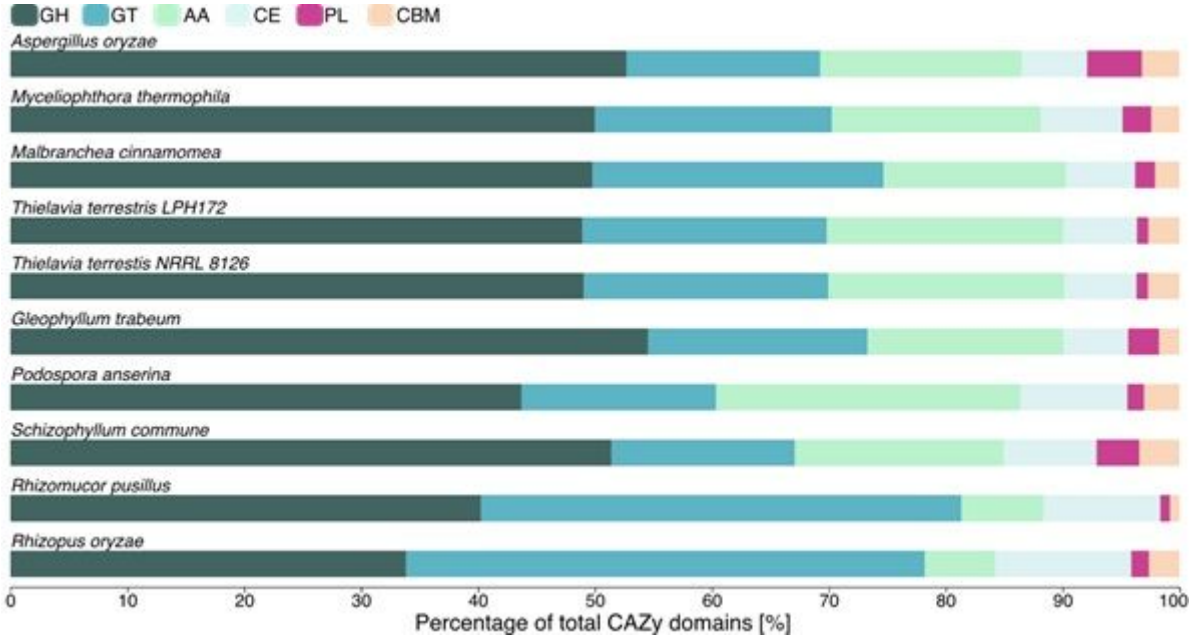

Figure 2

Relative numbers of CAZy domains in various filamentous fungi. For each species, the numbers of predicted CAZy domains were normalised to the total number of predicted CAZy 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 (HMMR algorithm).

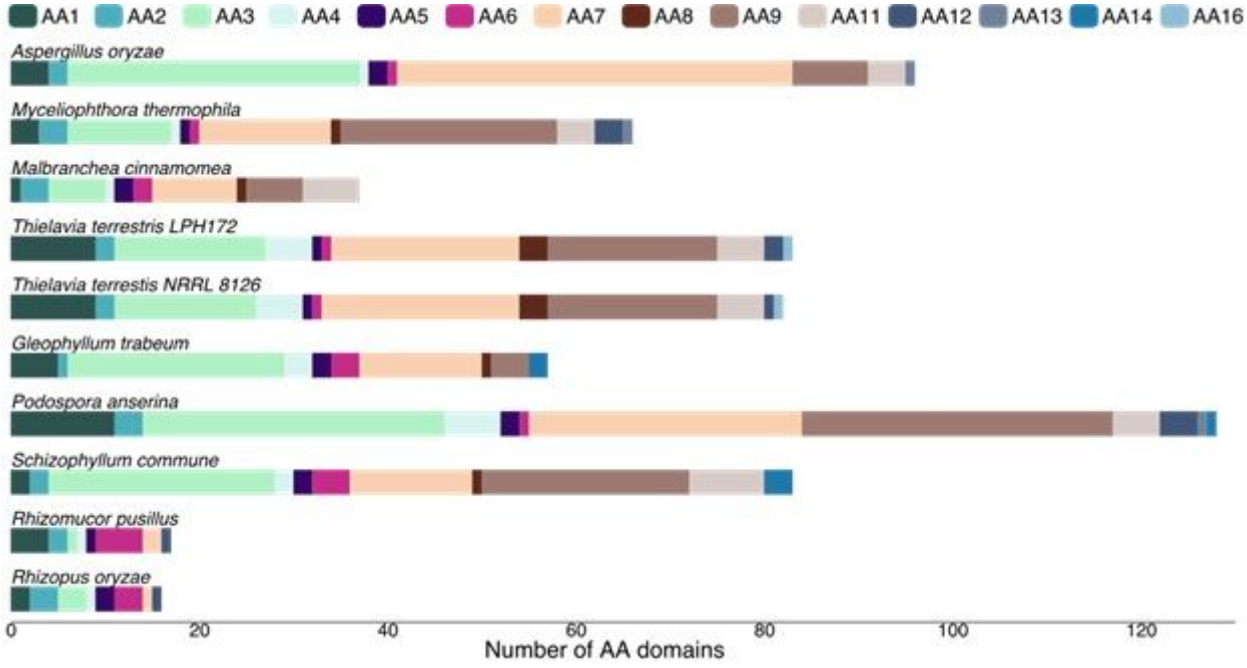

Figure 3

Auxiliary activity (AA) CAZyme subfamilies in different filamentous fungi. For each species, the number of predicted AA domains categorised into AA families 1–16 is shown. Predictions were made with dbCAN2 (HMMR algorithm).

Transcript ID	Predicted activity	SP	CAZy-domain(s)	fpkm				Upregulation			Putative substrate
				Avicel	RS	BX	Glc	Avicel	RS	BX	
TT_03917	Multicopper oxidase		AA1_2	1				15			L
TT_07374	Multicopper oxidase		AA1_2	5		8	1	4		6	L
TT_07941	Oxidoreductase	s	AA1_3	3		3	1	3		3	L
TT_00702	Laccase		AA1_3	34	2	1		3			L
TT_01663	Oxidoreductase		AA1_3	31	43	25	20		2		L
TT_02666	Oxidoreductase		AA3	29	17	24	3	6	5	6	C
TT_07514	Alkanesulfonate monooxygenase		AA3	10	10	2	2	4	6		C
TT_05013	Oxidoreductase	s	AA3	237	13	68	43	4			C
TT_04380	Cellobiose dehydrogenase	s	AA3_1-AA8	777	6	2	1	737	7		C
TT_09069	Cellobiose dehydrogenase	s	AA3_1-AA8	27	2	4	4	5			C
TT_08234	Oxygen-dependent choline dehydrogenase		AA3_2	170	4	78	3	37		17	L
TT_05138	Oxidoreductase	s	AA3_2	5				12			L
TT_05809	Oxidoreductase	s	AA3_2	3	1	1		8	2		L
TT_07512	Alcohol oxidase		AA3_3	1	3	1	2		2		L
TT_00225	Vanillyl-alcohol oxidase		AA4	1326	16	6	2	426	7		L
TT_06681	Oxidoreductase	s	AA7	35	13	2	1	20	10		
TT_02325	Cellobiose dehydrogenase	s	AA8		28				1168		C
TT_09190	Quinate permease		AA8	5	49	1	3		19		
TT_08370	LPMO	s	AA9	2417	262	7	1	1793	265	5	C
TT_07455	LPMO		AA9	1822	352	6	14	84	25		C
TT_06268	LPMO	s	AA9	5	60	3		64	1134	36	C
TT_04352	LPMO	s	AA9	88	224	1	1	43	150		C
TT_03770	LPMO		AA9	25	192	26	16		12		C
TT_01736	LPMO	s	AA9-CBM1	931	21	7		4644	140	36	C
TT_04350	LPMO	s	AA9-CBM1	926	6	3	1	468	4		C
TT_02354	LPMO		AA9-CBM1	2	1	1		5	3		C
TT_06499	Feruloyl esterase	s	CBM1	2110	27	4	3	525	9		X
TT_00910	LysM domain-containing protein		CBM50	38	46	1		152	256	2	Ch
TT_05636	Acetylxylan esterase	s	CE1	2	87				8411		X
TT_06012	Acetylesterase	s	CE16	501	128	5	2	147	52		X
TT_07814	Acetylesterase		CE16	93	47	11	9	8	5		X
TT_05762	Acetylxylan esterase	s	CE5	10	16	25		26	55	67	X
TT_08797	Cutinase	s	CE5	2	483	4	3		143		X
TT_08166	Acetylxylan esterase	s	CE5-CBM1	1450	95	554	20	54	5	21	X
TT_08613	Beta-glucosidase		GH1	1	18	1			745		C
TT_09033	Endo-1,4-β-xylanase	s	GH10	2	42	1		5	201	5	X
TT_08161	Endo-1,4-β-xylanase	s	GH10-CBM1	457	38	79	11	30	3	5	X
TT_09019	Rhamnogalacturonyl hydrolase		GH105		1	10	4			2	P
TT_01839	Endo-1,4-β-xylanase	s	GH11	213	2461	72	2	102	1610	35	X
TT_02489	Endo-1,4-β-xylanase	s	GH11	8	43	1		15	105		X
TT_03205	Endo-1,4-β-xylanase	s	GH11	2	24	1			152		X
TT_03075	Endo-1,4-β-xylanase	s	GH11-CBM1	2251	167	151	11	151	15	10	X
TT_05534	Endoglucanase		GH12	3	8	2	2		5		C
TT_01441	α-amylase	s	GH13_1-CBM20	91	290	103	45		6	2	S
TT_01042	Glycogen debranching enzyme		GH13_25	57	3	56	21	2		2	S
TT_00996	α-glucosidase		GH13_40	1		1		2		3	S
TT_08676	α-glucosidase		GH13_40	37	51	46	22		2		S
TT_01081	β-glucanase	s	GH131	314	224	21	14	16	15		
TT_02394	β-glucanase		GH131	1	3	3	1		4	5	
TT_02410	Glycosidase	s	GH16	82	9	87	20	3		3	C, XG
TT_04001	β-glucanase	s	GH16	8	1	5	3	2			C, XG
TT_08907	β-glucanase	s	GH16	16	25	63	10		3	5	C, XG
TT_03186	Chitinase		GH18	14		2		21			Ch
TT_00823	Chitinase		GH18	31	6	11	6	4			Ch
TT_05685	Chitinase		GH18	3	11	46	1	2	9	28	Ch
TT_00524	Chitinase		GH18	25	53	20	16		3		Ch
TT_04717	Chitinase	s	GH18-CBM18		2	8	1		3	11	Ch
TT_00912	Chitinase		GH18-CBM18-CBM50	49	61	4	5	7	13		Ch
TT_08320	Exo-β-glucosaminidase	s	GH2	1	1	1		2			GM
TT_00254	β-mannosidase		GH2	36	6	27	15	2			GM
TT_07949	β-mannosidase		GH2			1				4	GM
TT_09589	α-galactosidase		GH27	15	28	8	5	2	6		XG, X
TT_06031	Rhamnogalacturonase	s	GH28	55	13	79	3	14	6	11	P
TT_07738	Rhamnogalacturonase	s	GH28	21	3	24	7	2		2	P
TT_07104	Rhamnogalacturonase	s	GH28		3				36		P
TT_06264	Exopolygalacturonase	s	GH28	15	36	21	13		3		P
TT_01132	β-glucosidase		GH3	53	5	22	11	4			C
TT_09224	β-glucosidase		GH3	9	2	13	2	3		5	C
TT_03455	α-glucosidase	s	GH31	47	7	53	11	3		3	XG, S
TT_05828	α-xylosidase		GH31	9	8	3	5		2		XG
TT_05469	β-xylosidase		GH39	1	3	1	1		5		XG, X
TT_06379	Endo-α-1,5-L-arabinanase	s	GH43_24	40	2	17	2	13		6	XG, X, P
TT_00222	Arabinan endo-1,5-α-L-arabinosidase	s	GH43_30	4		1		18		3	XG, X, P
TT_02313	α-L-arabinofuranosidase		GH43_36	34	1	1	1	33			XG, X, P
TT_09000	Endo-β-1,4-glucanase	s	GH45	790	18	6	15	37			C
TT_01019	Endo-β-1,4-glucanase	s	GH5_5	679	26	60	10	50	3	5	C
TT_03238	Endo-β-1,4-glucanase	s	GH5_5	194	2	2	4	40			C
TT_05879	Endo-β-1,4-glucanase	s	GH5_5	75	27	66	17	3	2	3	C
TT_09640	Mannan endo-1,4-β-mannosidase	s	GH5_7	36	11	1	1	19	8		GM
TT_06537	Mannan endo-1,4-β-mannosidase		GH5_7	9	53				1951		GM
TT_06655	1,4-β-D-glucan cellobiohydrolase	s	GH6	2263	51	32	14	125	4	2	C
TT_09103	1,4-β-D-glucan cellobiohydrolase	s	GH6	18	3	19	5	3		3	C
TT_09005	α-L-arabinofuranosidase	s	GH62	1	19	1		18	583	14	XG, X, P
TT_05512	Endo-β-1,4-glucanase	s	GH7	122	3			5217			C
TT_03032	Endoglucanase	s	GH7	68	3	3	1	53			C
TT_04861	Exoglucanase	s	GH7	8	12	7	7		2		C
TT_05797	Endoglucanase	s	GH7-CBM1	3343	31	1	8	298	4		C
TT_05546	Exoglucanase	s	GH7-CBM1		34	10	11		3		C
TT_07085	Mannan endo-1,6-α-mannosidase	s	GH76	9	1	11	2	3		4	GM
TT_02335	Rhamnogalacturonate lyase	s	PL4_3	1	2	1			3		P

Figure 4

Expression and upregulation of CAZymes during cultivation on four different substrates. Putative CAZymes involved in biomass degradation were analysed for their expression levels (fpkm, fragments per

kilobase million), as well as their differential expression (Upregulation). Fpkm values show the normalised expression levels of transcripts (average of three replicates) on Avicel, rice straw (RS), beechwood xylan (BX), and glucose (Glc). Shading ranges from low expression (light blue) to high expression (magenta). Blank cells indicate expression levels of fpkm < 1. Upregulation shows the fold-change in gene expression during cultivation on Avicel, RS, and BX compared to cultivation on glucose. Shading of upregulated genes (i.e. 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 upregulation was not significant are indicated by blank cells. Transcripts were considered significantly differentially expressed when fold-change was at least 2 ($p \leq 0.05$). In a few cases of very high, but according to the criteria non-significant upregulation, the numbers were still included in the figure (in italics and white). All numbers were rounded to the nearest integer, which is why some genes show the same fold-change despite different fpkm numbers. The Predicted Activity of the gene products is based on BLASTp predictions. CAZy domains were analysed with dbCAN2. The presence of putative signal peptides (SP), predicted by SignalP 4.0, is indicated by a small, blue s. Putative substrates of the CAZymes are: C, cellulose; Ch, chitin; GM, glucomannan; L, lignin; P, pectin; S, starch; X, xylan; XG, xyloglucan.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1final.docx
- Additionalfile2final.docx
- Additionalfile3final.xlsx
- Additionalfile4updated1.xlsx
- Additionalfile5final.xlsx