The synergistic actions of hydrolytic genes in coexpression networks reveal the potential of *Trichoderma harzianum* for cellulose degradation

Déborah Aires Almeida[1,2], Maria Augusta Crivelente Horta[1,2,#], Jaire Alves Ferreira Filho[1,2], Natália Faraj Murad[1] and Anete Pereira de Souza[1,3,*]

[1]Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil

[2]Graduate Program in Genetics and Molecular Biology, Institute of Biology, UNICAMP, Campinas, SP, Brazil

[3]Department of Plant Biology, Institute of Biology, UNICAMP, Campinas, SP, Brazil

[#] Present Address: Holzforshung München, TUM School of Life Sciences Weihenstephan, Technische Universität München, Freising, Germany

***Corresponding author**

Profa Anete Pereira de Souza

Dept. de Biologia Vegetal, Universidade Estadual de Campinas, CEP 13083-875, Campinas, São Paulo, Brazil

Tel.: +55-19-3521-1132

E-mail:anete@unicamp.br

Abstract

Background: Bioprospecting key genes and proteins related to plant biomass degradation is an attractive approach for the identification of target genes for biotechnological purposes,

especially genes with potential applications in the biorefinery industry that can enhance second-generation ethanol production technology. *Trichoderma harzianum* is a potential candidate for cellulolytic enzyme production. Herein, the transcriptome, exoproteome, enzymatic activities of extracts, and coexpression networks of the *T. harzianum* strain CBMAI-0179 under biomass degradation conditions were examined.

Results: We used RNA-Seq to identify differentially expressed genes (DEGs) and carbohydrate-active enzyme (CAZyme) genes related to plant biomass degradation and compared them with genes of strains from congeneric species (*T. harzianum* IOC-3844 and *T. atroviride* CBMAI-0020). *T. harzianum* CBMAI-0179 harbors species- and treatment-specific CAZyme genes, transporters and transcription factors. Additionally, we detected important proteins related to biomass degradation, including β-glucosidases, endoglucanases, cellobiohydrolases, lytic polysaccharide monooxygenases (LPMOs), endo-1,4-β-xylanases and β-mannanases, in the exoproteome under cellulose growth conditions. Coexpression networks were constructed to explore the relationships among the genes with corresponding secreted proteins that act synergistically for cellulose degradation. An enriched cluster with degradative enzymes was described, and the subnetwork of CAZymes showed linear correlations among secreted proteins (AA9, GH6, GH10, GH11 and CBM1) and differentially expressed CAZyme genes (GH45, GH7, AA7 and GH1).

Conclusions: The coexpression network revealed genes with strong correlations acting synergistically to hydrolyze cellulose. Our results provide valuable information for future studies on the genetic regulation of plant cell wall-degrading enzymes. This knowledge can be exploited for the improvement of enzymatic reactions to degrade plant biomass, which is useful for bioethanol production.

Keywords: *Trichoderma harzianum*, Cellulose, RNA-Seq, CAZymes, Exoproteome, Coexpression networks

Background

The expanding worldwide demand for renewable and sustainable energy sources has increased the interest in alternative energy sources, and the production of second-generation biofuels seems to be the most viable option to confront these issues [1, 2]. Lignocellulosic biomass is the most abundant renewable organic carbon resource on earth, consisting of three major polymers, cellulose, hemicellulose, and lignin [2]. However, due to its recalcitrant characteristics that prevent enzyme access, degrading this complex matrix is still a major challenge [3]. For the complete hydrolysis of lignocellulose, a variety of enzymes acting in synergy are required, and much research has focused on this topic in recent decades [4]. Interactions between different enzymes have been investigated to identify optimal combinations and ratios of enzymes for efficient biomass degradation, which are highly dependent on the properties of the lignocellulosic substrates and the surface structure of cellulose microfibrils [4, 5]. Due to their abundance in nature, microorganisms are considered natural producers of enzymes, and many of them, including members of both bacteria and fungi, have evolved to digest lignocellulose [6, 7]. The search for microorganisms that are able to efficiently degrade lignocellulosic biomass is pivotal for the establishment of the sustainable production of bioethanol [8].

Filamentous fungi, including the genera *Trichoderma*, *Aspergillus*, *Penicillium* and *Neurospora*, produce extracellular proteins that act synergistically to degrade plant cell walls and are widely used in the enzymatic industry [9]. Species in the filamentous ascomycete genus *Trichoderma* are among the most commonly isolated saprotrophic fungi [10] and are important from a biotechnological perspective [7]. *Trichoderma* species are widely used in

agriculture as biocontrol agents due their ability to antagonize plant-pathogenic fungi and in industry as producers of plant cell wall-degrading enzymes [11-14]. In addition, *Trichoderma* species are easily isolated from soil and decomposing organic matter [15]. Within the *Trichoderma* genus, *T. reesei* is the most intensively studied species [16]. *T. reesei* is a well-known producer of cellulase and hemicellulase, and due to the high effectiveness of the synergistic cellulases in this species, it is widely employed in industry, as technologies for its use and handling are based on seventy years of experience [8, 16-19]. However, studies on *T. harzianum* strains have shown their potential to produce a set of enzymes that can degrade lignocellulosic biomass [20-23]; therefore, *T. harzianum* strains are being investigated as potentially valuable sources of industrial cellulases [6].

The identification of carbohydrate-active enzymes (CAZymes) that act synergistically under biodegradation conditions [4] has the potential to improve the enzymatic hydrolysis process by optimizing and reducing bioethanol costs. The CAZy database (www.cazy.org) classifies CAZymes into six major groups: glycoside hydrolases (GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AAs), and carbohydrate-binding modules (CBMs) [24]. CAZymes are extensively used for the genetic classification of important hydrolytic enzymes [22, 25].

The conversion of cellulose to glucose involves the synergistic action of three principal groups of enzymes: endo-β-1,4-glucanases (EC 3.2.1.4), β-glucosidases (EC 3.2.1.21), and cellobiohydrolases (EC 3.2.1.91/176) [20, 26]. For hemicellulose hydrolysis, several enzymes are needed, such as endo-1,4-β-xylanases (EC 3.2.1.8), β-xylosidases (EC 3.2.1.37), β-mannanases (EC 3.2.1.78), arabinofuranosidases (EC 3.2.1.55), and acetylxylan esterases (EC 3.1.1.72) [26, 27]. In addition, a number of auxiliary enzymes are involved in this process, such as lytic polysaccharide monooxygenases (LPMOs), cellulose-induced

protein 1 and 2 (CIP1 and CIP2) and swollenin, which can increase the hydrolytic performance of enzymatic cocktails used in industry for bioethanol production [6, 28-30].

As genetic variation occurs within species [31, 32], understanding and exploring the genetic mechanisms of different *T. harzianum* strains can provide valuable information for industrial applications. In the present study, we analyzed the enzymatic activity, transcriptome and exoproteome of *T. harzianum* CBMAI-0179 and compared them with those of other *Trichoderma* strains (*T. harzianum* IOC-3844 and *T. atroviride* CBMAI-0020). RNA-Seq analysis was performed to construct coexpression networks, providing novel information about the potential of this *T. harzianum* strain for biotechnological applications. Our findings provide insights into the genes/proteins that act synergistically in plant biomass conversion and can be exploited to improve enzymatic hydrolysis and thereby increase the efficiency of the saccharification of lignocellulosic substrates for bioethanol production.

Results

Transcriptome analysis of Trichoderma *spp. under cellulose growth conditions*

The present study represents the first deep genetic analysis of Th0179, describing and comparing the transcriptome by RNA-Seq under two different growth conditions, cellulose and glucose, to identify the genes involved in plant biomass degradation. Reads were mapped against reference genomes of *T. harzianum* (PRJNA252551) and *T. atroviride* (PRJNA19867), generating 96.3, 111.8 and 133.3 million paired-end reads for Th0179, Th3844 and Ta0020, respectively. To establish the degrees of similarity and difference in gene expression among strains and between treatments, a principal component analysis (PCA) was performed using the *T. harzianum* T6776 genome as a reference. The PCA results

showed clustered groups with higher similarity between treatments than among strains. The transcriptomes of Th0179 and Th3844 were more similar to each other than to the transcriptome of Ta0020 (Fig. 1a), showing that it is possible to capture differences among the three strains.

Venn diagrams of the genes exhibiting expression levels greater than zero were constructed based on the similarities among the genes from all strains, thus showing the species-specific genes expressed under both conditions (Additional file 1: Fig. S1). Through the transcriptome analysis of the strains, we identified 11,250 genes exhibiting expression levels greater than zero under cellulose growth conditions and 11,235 genes exhibiting expression levels greater than zero under glucose growth conditions. The number of genes shared by Th0179 and Th3844 was higher under both conditions than that shared by either Th0179 or Th3844 and Ta0020. Th0179 exhibited the highest number of unique expressed genes, with 374 and 168 unique genes under the cellulose and glucose growth conditions, respectively. Among these unique genes under cellulose growth conditions, we found major facilitator superfamily (MFS) transporters (THAR02_00234, THAR02_00911, THAR02_03251, THAR02_03935, THAR02_07021, THAR02_07705 and THAR02_07942), an ATP-binding cassette (ABC) transporter (THAR02_09958), a drug resistance protein (THAR02_04837), a fungal specific transcription factor (TF) (THAR02_07743), and a C2H2 TF (THAR02_11070).

Among the genes that were upregulated in cellulose conditions relative to glucose conditions, 219 were identified for Th0179, 281 were identified for Th3844, and 718 were identified for Ta0020 (Fig. 1b and Additional file 2: Table S1). We validated the in silico analyzes using a subset of the DEGs under cellulose or glucose growth conditions for all strains through an independent technique, i.e., RT-qPCR (Additional file 3: Fig. S2).

CAZyme identification and distribution in* Trichoderma *spp.

For CAZymes important in the degradation of biomass, the set of CAZymes was identified by mapping all of the proteins of *T. harzianum* T6776 and *T. atroviride* IMI 206040 against the CAZy database using the BLASTp search tool. Based on the filtering criteria, a total of 631 proteins were retained as CAZyme genes for *T. harzianum*, which corresponds to 5.5% of the total 11,498 proteins predicted for this organism [33], and 640 proteins were retained for *T. atroviride*, which corresponds to 5.4% of the total 11,816 proteins predicted for this organism [34].

Considering the DEGs under cellulose growth conditions and using the established CAZy database, we found 35, 78 and 31 differentially expressed CAZyme genes in our data (Fig. 1b) for Th0179, Th3844 and Ta0020, respectively. We identified the main CAZyme classes (AA, GH, GT, CE and CBM) and their contents for all strains (Fig. 1c). The Th3844 strain presented the highest number of classified genes from the AA family, a high number of CBMs and twice the number of identified GHs found in the two other strains. Strain Ta0020 presented a higher number of genes from the GT family than Th0179, which exhibited a higher number of classified genes from the CBM family than Ta0020. The differences regarding the specific CAZyme families for each strain are shown in Fig. 2.

The GH group was the most represented class of enzymes present in all evaluated strains. GHs are key enzymes for carbohydrate hydrolysis and include enzymes capable of degrading cellulose [35-37], with many able to cleave glycosidic bonds between glucose molecules. Another important family involved in degradation of the plant cell wall, the AA family, was also identified in all strains [38]. The AA class currently harbors 9 families of ligninolytic enzymes and 6 families of lytic polysaccharide mono-oxygenases that may not act on carbohydrates. However, because lignin is invariably and intimately associated with carbohydrates in the plant cell wall, these lignolytic enzymes cooperate with classical

polysaccharide depolymerases, named auxiliary enzymes [24, 38], that differentially influence degradative activity among species.

Cellulose and hemicellulose degradative enzymes in* Trichoderma *spp.

In plant biomass degradation, a variety of enzymes working synergistically are required for complete hydrolysis [4]. Under cellulose growth conditions, we found upregulated families among the CAZymes responsible for cellulose degradation, including endoglucanases (GH5, GH12, GH45 and CBM1), β-glucosidases (GH1 and GH3) and cellobiohydrolases (GH6 and GH7), that play important roles in cellulose degradation. In addition, among the CAZymes responsible for hemicellulose degradation, we identified endo-1,4-β-xylanases (GH10, GH11 and CBM1), arabinofuranosidases (GH54, GH62 and CBM42) and acetylxylan esterases (CE5 and CBM1). Additionally, we found the copper enzymes LPMOs, classified as AAs in the family AA9 [39]; these are considered a breakthrough in the enzymatic degradation of cellulose because they oxidatively cleave glycosidic linkages that render substrates more susceptible to hydrolysis by conventional cellulases [28]. The expression levels of the main cellulase and hemicellulase families based on their principal enzyme activity present in Th0179, Th3844 and Ta0020 were evaluated using the transcriptomic data (Fig. 3).

All of the GHs related to cellulose degradation were found in Th3844, including 2 AA9 members with a CBM1 module for cellulose binding. All GHs except GH12, including AA9/CBM1, were detected in Th0179. Only one gene belonging to the GH5 family (TRIATDRAFT_81867) with cellulase activity (EC 3.2.1.4) was observed in Ta0020 (48.12 TPM). The most highly expressed genes were cellobiohydrolases from the GH6/CBM1 (THAR02_04414 – 299.66 TPM for Th0179 and 986.25 TPM for Th3844) and GH7/CBM1

families (THAR02_08897 – 369.92 TPM for Th0179; THAR02_03357 and THAR02_08897 – 1876.89 TPM for Th3844) (Fig. 3a).

In addition, 5 families related to hemicellulose degradation were identified in Th3844, in which endo-1,4-β-xylanases (EC 3.2.1.8) from the GH11 family had the greatest expression levels (THAR02_02147, THAR02_05896, THAR02_08630 and THAR02_08858 – 1013.62 TPM). Endo-1,4-β-xylanase (EC 3.2.1.8) from the GH10 family was found in both *T. harzianum* strains (THAR02_03271 – 47.58 TPM for Th0179 and 741.35 TPM for Th3844). Acetylxylan esterases (EC 3.1.1.72) from the CE5/CBM1 family were found only in Th3844 (THAR02_01449 and THAR02_07663 – 227.56 TPM). We identified only one α-L-arabinofuranosidase (EC 3.2.1.55) from the GH54/CBM42 family (TRIATDRAFT_81098 – 18.07 TPM) in Ta0020, whereas for Th3844, three α-L-arabinofuranosidases from the GH54/CBM42 and GH62 families were identified (Fig. 3b). The Ta0020 strain showed the lowest number of genes and the lowest expression levels of the CAZyme families related to cellulose and hemicellulose degradation. The classification of the CAZyme genes along with their enzyme activities, fold change values, e-values, EC numbers and expression values (TPM) for Th0179 under cellulose fermentative conditions are described in Table 1. The corresponding information for the Th3844 and Ta0020 strains can be found in Additional file 4: Table S2.

[Insert Table 1 here]

Functional annotation of* T. harzianum *CBMAI-0179 in the presence of cellulose

The first functional annotation of the expressed genes under cellulose fermentative conditions for Th0179 was performed based on GO terms (Fig. 4). A total of 7,718 genes were annotated, which corresponds to 67.1% of the total 11,498 genes predicted for this organism (*T. harzianum* T6776) [33]. Under the molecular function category, catalytic

activity was the main annotated function, with 3,350 identified genes. Other functions that were enriched under this category were hydrolase activity, nucleotide binding and TF activity, with 1,111, 899 and 277 genes, respectively. For the biological process category, the main functions were metabolic process, with 3,607 genes, and cellular process, with 2,575 genes. In addition, within the biological process category, 574 genes were identified as associated with transmembrane transport, and 66 genes were identified as associated with regulation of catalytic activity. Compared with Th3844, Th0179 presented 163 more genes related to catalytic activity, 53 more genes related to hydrolase activity, and 15 more genes associated with TF activity but 10 fewer genes related to transmembrane transport, suggesting that these two strains have developed different functional regulation. Ta0020 presented fewer genes related to any of the above functions than Th3844 and Th0179 except for TF activity, for which Ta0020 had 53 more genes than Th0179.

Exoproteome and RNA-Seq data correlation of* T. harzianum *CBMAI-0179

Once the transcriptome was characterized, we analyzed the secreted proteins identified in the exoproteome profile of *T. harzianum* CBMAI-0179. A total of 64 proteins, which had been secreted and were present in the culture medium after 96 h of fermentation, were detected in the extracts. Of those, 32 proteins were present in the cellulose aqueous extract (Table 2), 12 were present only in the glucose aqueous extract, and 20 were present under both conditions (Additional file 5: Table S3). Among the 32 secreted proteins detected using cellulose as the carbon source, the main CAZyme families observed exclusively in this supernatant were among of the most important families for cellulose and hemicellulose degradation, such as β-glucosidases (EC 3.2.1.21), endo-β-1,4-glucanases (EC 3.2.1.4), LPMOs (EC 3.2.1.4), cellobiohydrolases (EC 3.2.1.91), endo-1,4-β-xylanases (EC 3.2.1.8), and β-mannanases (EC 3.2.1.78). Two genes from the GH3 (THAR02_00656 and

THAR02_00890) and GH5/CBM1 (THAR02_04405 and THAR02_09719) families and one gene from the AA9/CBM1 (THAR02_02134) and GH6/CBM1 (THAR02_04414) families corresponded to the main group of secreted cellulases, whereas one gene from the GH10 (THAR02_03271) family and two genes from the GH11 (THAR02_02147 and THAR02_05896) family corresponded to the main group of hemicellulases detected in the supernatant. In addition, a member of the GH5 cellulase family with β-mannanase activity (THAR02_03851) was identified. Besides these secreted proteins, 8 other proteins were classified as uncharacterized proteins. We also identified hemicellulose-degrading enzymes in both extracts, such as α-L-arabinofuranosidase B (EC 3.2.1.55) and xylan 1,4-β-xylosidase (EC 3.2.1.37) (Additional file 5: Table S3).

[Insert Table 2 here]

Correlating the exoproteome data with the transcriptome data under cellulose growth conditions, we observed the expression levels of genes that play important roles in plant biomass degradation. CAZyme genes showing high TPM values in cellulose included cellulose 1,4-β-cellobiosidase (nonreducing end) (EC 3.2.1.91) from the GH6/CBM1 family (299.66 TPM), cellulase (EC 3.2.1.4) from the AA9/CBM1 family (138.62 TPM) and endo-1,4-β-xylanase (EC 3.2.1.8) from the GH11/CBM1 family (119.56 TPM). Two uncharacterized proteins (THAR02_02133 and THAR02_08479) with the CBM1 module of cellulose binding also showed increased expression levels under the cellulose condition. In contrast, the THAR02_00656 gene, which displays β-glucosidase (EC 3.2.1.21) activity from the GH3 family, had the lowest expression level (5.08 TPM) among the CAZyme genes related to biomass degradation, indicating that genes with low expression levels are also important for functional secreted proteins [40].

Coexpression networks for* T. harzianum *CBMAI-0179

The coexpression network was assembled for Th0179 and included all genes obtained from the mapping results using the *T. harzianum* T6776 genome as a reference under both conditions. The DEGs and CAZyme genes were highlighted in the network (Fig. 5a). This network, constructed based on the expression level data, was composed of a total of 11,104 nodes with 153,893 edges. We detected 219 genes corresponding to the DEGs under cellulose growth conditions and 367 genes corresponding to the nodes under glucose growth conditions. The respective CAZyme genes from both conditions were also identified in this network, with 35 CAZymes under cellulose growth conditions and 28 CAZymes under glucose growth conditions. The DEGs tended to cluster in the top (DEGs in glucose) and bottom (DEGs in cellulose) parts of the network, reflecting the different regulation of degradative activities according to substrate.

A subnetwork was generated based only on the secreted proteins that were present in the cellulose aqueous extract and that had corresponding genes in the coexpression network (Fig. 5b). This subnetwork exclusively represented the genes and their closest related genes that are correlated to the proteins secreted. It was composed of 713 nodes and 6,124 edges, including the 32 genes that encode the secreted proteins. In this subnetwork was also identified 39 DEGs under cellulose growth conditions, 8 DEGs under glucose growth conditions, 6 CAZyme genes related to cellulose degradation and 1 CAZyme gene under glucose growth conditions. Among the CAZyme genes under cellulose growth conditions, the GH1 family (THAR02_02251 and THAR02_05432) with β-glucosidase activity, the GH7/CBM1 family (THAR02_08897) with cellulose 1,4-β-cellobiosidase activity, and the GH45/CBM1 family (THAR02_02979) with cellulase activity were found in this subnetwork. Despite the different functions of the related genes, it is predicted that these genes participate in the genetic regulation of the detected CAZymes/proteins and are important to the regulation of the hydrolytic system.

The cluster analysis classified 11,102 genes in 84 clusters (Additional file 6: Table S4). We identified an enriched cluster composed of 196 nodes and 2,125 edges (Fig. 5c and Additional file 6: Table S4) containing the greatest number of CAZyme genes among the clusters. Of the 12 CAZyme genes detected, 7 corresponded to the secreted proteins detected in the cellulose aqueous extract, and 5 were differentially expressed. In addition, we identified 20 DEGs under cellulose growth conditions and 119 uncharacterized proteins in this cluster, showing a strong correlation with the CAZyme genes, which indicates that the unknown genes are important to the degradation process. The CAZyme genes from the cluster analysis were selected to generate a new subnetwork with their corresponding edges (Fig. 5d), showing linear correlations among secreted proteins (AA9, GH6, GH10, GH11 and CBM1) and differentially expressed CAZyme genes (GH45, GH7, AA7 and GH1).

Discussion

In this study, different biotechnological approaches were used in bioprospecting new and efficient enzymes for possible applications in the enzymatic hydrolysis process. We performed enzymatic activity, transcriptome, exoproteome and coexpression network analyses of the *T. harzianum* strain CBMAI-0179 that has potential for plant biomass degradation under cellulose growth conditions to gain insights into the genes and proteins produced, associated with cellulose hydrolysis. The analyses of genetic expression together with the identified secreted proteins under biomass degradation conditions allowed us to construct coexpression networks to investigate the relationships among the genes. Furthermore, we compared the expression levels with *Trichoderma* spp. that have been previously studied under the same conditions [10].

Several studies, including transcriptomic and proteomic studies, have been performed using filamentous fungi to bioprospect efficient catalysts for the development and

improvement of enzymatic cocktails to degrade plant biomass. Some fungi, including *T. reesei*, *A. niger*, *A. nidulans* and *N. crassa* [9, 16, 17, 41-43], have been used for this purpose; however, the search for new efficient strains is ongoing. Understanding the molecular mechanisms by which filamentous fungi degrade plant biomass can improve the saccharification process, a very important step in the production of second-generation ethanol [42, 44]. Beyond that, the identification of new species and powerful enzymes can enhance the technologies in the biofuel industry.

Trichoderma spp. are capable of degrading both fungal and plant cell wall materials [45]. In this study, we chose to investigate species of the genus *Trichoderma* because this genus is a common soil and wood-degrading fungi distributed worldwide [46-48], is easily isolated from decomposing organic matter and soil [49], and harbors great potential to produce enzymes that degrade plant biomass [10, 12, 22]. We selected two strains of *T. harzianum* (Th0179 and Th3844) and one strain of *T. atroviride* (Ta0020) to capture differences among strains.

Through the transcriptome analysis, we identified the DEGs in all strains (Fig. 1b), and even though Ta0020 presented the highest number of identified DEGs, this strain is less efficient than other *Trichoderma* strains at degrading plant biomass. *T. atroviride* is mostly used as a biocontrol agent and is among the best mycoparasitic fungus used in agriculture [34, 50, 51]. Although the *T. atroviride* strain produced 31 differentially expressed CAZymes, they are less efficient enzymes for plant biomass degradation with lower expression levels than *T. harzianum*, since it was observed to harbor only one gene with cellulase activity from the GH5 family (TRIATDRAFT_81867) and only one gene with hemicellulase activity from the GH54/CBM42 family (TRIATDRAFT_81098) (Fig. 3). In this analysis, each strain presented a different set of genes with different expression levels, which can be attributed to strain differences in the regulatory mechanisms of hydrolysis.

Additionally, the carbon source, which plays an important role in the production of enzymes, promotes the expression of different sets of genes as the fungus seeks to adapt to new environments [10]. The *T. harzianum* strains showed high numbers of CAZyme genes with enhanced specificity for biomass degradation.

Among the main CAZyme classes detected in all strains, GHs were well represented, with 21 genes in Th0179 and twice that in Th3844 (Fig. 1c). GHs compose an extremely important class in several metabolic routes in fungi, including genes involved in cellulose and chitin degradation [12, 36]. Here, we found the main CAZyme families responsible for cellulose degradation, such as GH1, GH3, GH5, GH6, GH7, GH12, GH45 and an LPMO from the AA9 family [12, 28]. The CAZyme families responsible for hemicellulose degradation were as follows: GH10, GH11, GH54, GH62 and CE5 [12, 30, 52-54]. Within the GH class, an important family expected in our data was GH18, which has chitinase activity; this family was expected since members of the genus *Trichoderma* (such as *T. harzianum* and *T. atroviride*) are capable of mycoparasitism and because this class is directly related to the biological control of these species [17, 35, 55].

In comparing the two *T. harzianum* strains (Th0179 and Th3844), we observed that Th0179 presented fewer CAZyme genes related to cellulose degradation than Th3844, with lower expression levels (Fig. 3a). However, in measuring the enzymatic activity from the culture supernatants after 96 h of growth, we found that both strains had similar cellulase activity profiles during growth on cellulose (Additional file 7: Fig. S3), suggesting greater potential of Th0179 to degrade cellulose. The detected enzymatic activity is related to proteins secreted into the medium by the cells, and only the most stable proteins are detected in this environment [10]. It is interesting to observe which proteins found in the exoproteome of Th0179 may respond to this increased cellulase activity. A similar profile was observed in *T. reesei*, the most studied fungus within this genus and an important industrial producer of

cellulolytic enzymes [13]: few CAZyme genes have been detected in its machinery, but it can reach the highest cellulolytic activity [10, 19, 56].

The identification of LPMOs as a group of enzymes that accelerate the breakdown of carbohydrate polymers, such as cellulose, by oxidative cleavage has been a breakthrough in lignocellulose conversion research. The supplementation of AA9 enzymes in commercial cocktails improves the hydrolysis of lignocellulose; they assist cellulases in attacking crystalline substrate areas, resulting in rapid and relatively complete surface degradation [19].

CAZyme genes related to cellulose degradation, including endo-β-1,4-glucanases, β-glucosidases, and cellobiohydrolases, and CAZyme genes related to hemicellulose degradation, such as endo-1,4-β-xylanases, were identified in the Th0179 transcriptome under cellulose growth conditions (Table 1). In addition, we verified important secreted CAZymes that play important roles in biomass degradation (Table 2), such as β-glucosidases (A0A0F9XRC5 and A0A0F9XQT4), LPMO (A0A0F9XMI8), cellulose 1,4-β-cellobiosidase (nonreducing end) (A0A0G0AEM7), cellulases (A0A0F9XG06 and A0A0F9WYH5), endo-1,4-β-xylanases (A0A0F9Y0Y9, A0A0H3UCP8, and A0A0F9XXA4), and endo-1,4-β-mannosidase (A0A0G0AGG8). One β-glucosidase from the GH3 family (THAR02_00656) was found in the exoproteome, which had the lowest expression level among the CAZymes related to biomass degradation; the same pattern was observed for *T. harzianum* IOC-3844 [10]. All of these enzymes are well known and can be used to improve enzymatic cocktails optimized for the degradation of specific substrates [57], such as the Brazilian biomass sugarcane bagasse. For instance, it is known that β-glucosidases are the rate-limiting enzymes in the degradation of cellulose [58]. These enzymes play a critical role among enzymes in enzymatic cocktails for biomass degradation. Therefore, improving the activities of such enzymes can enhance the efficiency of commercial enzymatic cocktails for bioethanol production [3].

Horta et al. [10] analyzed the exoproteome of Th3844 and Ta0020 under equivalent conditions, selecting a set of 80 proteins for a complete classification and analysis of expression levels based on their transcriptomes. Comparison of the Th0179 and Th3844 exoproteomes indicated that both produced many of the main CAZymes listed above; however, the THAR02_05896 gene, which encodes an endo-1,4-β-xylanase protein, and the THAR02_03851 gene, which encodes a protein with endo-1,4-β-mannosidase activity, were not detected in the Th3844 secretome. In the present study, when these exoproteomes were compared, we noticed that most of the proteins were upregulated in cellulose only for Th0179, with an emphasis on the AA9/CBM1 family (THAR02_02134), showing a 2.65-fold change (138.62 TPM). Among the strains, Ta0020 exhibited the lowest number of secreted proteins related to cellulose and hemicellulose degradation. Thus, the exoproteome analysis identified key enzymes that are fundamental for cellulose hydrolysis and act synergistically for efficient plant biomass degradation.

The organization of the transcriptomic data into coexpression networks using graph theory allowed the construction of gene interaction networks that were represented by nodes connected to edges [59]. Nodes represent the genes, and the edges represent the connections among these genes. Correlations are determined based on the expression level of the genes pair by pair, indicating that genes spatially closer to one another are more highly correlated than the genes that are farther apart. The coexpression subnetwork (Fig. 5b) revealed complex, specific relationships between CAZyme genes and genes involved in the production and secretion of the detected proteins and is helpful for understanding the functions and regulation of genes. Networks such as this one can be used as a platform to search for target genes or proteins in future studies to comprehend the synergistic relationships between genes, their regulation and protein production, which is very useful information for understanding

the saccharification process. We identified an enriched cluster with strong correlations among genes, CAZyme genes, secreted proteins and unknown genes in the cluster analysis.

Furthermore, through the functional annotation analysis of Th0179, we identified TFs and transporters, which should be investigated in further studies to better understand the mechanisms by which these genes are regulated. Most sugar transporters have yet to be characterized [1] but play important roles in taking up mono- or disaccharides into fungal cells after biomass degradation [1, 60]. Fungal gene expression is controlled at the transcriptional level [61], and its regulation affects the composition of enzyme mixtures; accordingly, it is explored in several species because of its potential applications [62]. This further emphasizes the importance of deleting and/or overexpressing TFs that regulate specific genes directly involved in plant biomass degradation [61].

In summary, the analyses of the enzymatic activity of cellulase, the transcriptome, the exoproteome and the coexpression networks revealed important enzymes that *T. harzianum* CBMAI-0179 uses to hydrolyze cellulose and that most likely act synergistically to depolymerize polysaccharides. The results suggest great potential of this strain to degrade cellulose and can contribute to the optimization of enzymatic cocktails for bioethanol production.

Conclusions

Bioprospecting new catalytic enzymes and improving technologies for the efficient enzymatic conversion of plant biomass are required for advancing biofuel production. *T. harzianum* CBMAI-0179 is a novel potential candidate producer of plant cell wall polysaccharide-degrading enzymes that can be biotechnologically exploited for plant biomass degradation. The cellulase activity profile indicated high efficiency and the potential of this strain for cellulose degradation. A set of highly expressed CAZymes and proteins that are

species- and treatment-specific was observed in both the transcriptome and exoproteome analyses. The coexpression network revealed coexpressed genes and CAZymes that act synergistically to hydrolyze cellulose. In addition, the cluster analysis revealed genes with strong correlations that are necessary for saccharification. Combined, these tools provide a powerful approach for catalysts discovery and the selection of target genes to the heterologous expression of proteins. In future studies, these tools can aid the selection of new species and the optimization of the production of powerful enzymes for use in enzymatic cocktails for second-generation bioethanol production.

Methods

Fungal strains, fermentation and enzymatic activities

The species originated from the Brazilian Collection of Environment and Industry Microorganisms (CBMAI), located on CPQBA/UNICAMP, in Campinas, Brazil. *T. harzianum* CBMAI-0179 (Th0179), *T. harzianum* IOC-3844 (Th3844) and *T. atroviride* CBMAI-0020 (Ta0020) strains were grown as described in a previous work on solid medium to produce sufficient spores for the fermentation process, which was performed in biological triplicates using crystalline cellulose (Celuflok, São Paulo, Brazil; degree of crystallinity, 0.72 g/g; composition, 0.857 g/g cellulose and 0.146 g/g hemicellulose) or glucose as the carbon source [10]. Glucose was used as a control in all experimental conditions. Supernatants were collected to measure enzymatic activity and determine the exoproteome profile.

Xylanase and β-glucosidase activities were determined using the methods described by Bailey and Poutanen [63] and Zhang et al. [64], respectively. Cellulase activity was determined using the filter paper activity (FPA) test according to Ghose [65]. Protein levels

were measured based on the Bradford [66] method. The enzymatic activities and protein contents in culture supernatants were determined in a previous work and are shown in Additional file 7: Fig. S3.

RNA extraction

Mycelial samples from cellulose and glucose conditions were extracted from Th0179, Th3844 and Ta0020 after 96 h of fermentation, stored at -80 °C, ground in liquid nitrogen using a mortar and pestle, and used for RNA extraction using the LiCl RNA extraction protocol according to the method reported by Oliveira et al. [67].

Library construction and RNA-Seq

RNA samples were quantified using a NanoDrop 8000 (Thermo Scientific, Wilmington, DE, USA). The libraries were constructed using 1 µg of each RNA sample obtained from the mycelial samples and the TruSeq RNA Sample Preparation Kit v2 [68] (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions. The expected target sizes were confirmed using a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and the DNA 1000 Kit, and the libraries were quantified by qPCR using the KAPA library quantification Kit for Illumina platforms (Kapa Biosystems, Wilmington, MA, USA). The average insertion size was 260 bp. A total of 18 biological triplicate samples were multiplexed with different adapters and organized in different lanes of the flow cell for high-throughput sequencing. The sequencing was carried out on the HiSeq 2500 platform (Illumina, San Diego, CA, USA) according to the manufacturer's specifications for paired-end reads of 150 bp.

Data sources

The reads were deposited into the SRA database in NCBI under BioProject number PRJNA336221 and accession numbers SAMN12583041, SAMN12583042, SAMN12583043, SAMN12583044, SAMN12583045, and SAMN12583046 for Th0179; SAMN12583047, SAMN12583048, SAMN12583049, SAMN12583050, SAMN12583051, and SAMN12583052 for Th3844; SAMN12583053, SAMN12583054, SAMN12583055, SAMN12583056, SAMN12583057, and SAMN12583058 for Ta0020. The nucleotide and protein sequences of *T. harzianum* T6776 (PRJNA252551) and *T. atroviride* IMI 206040 (PRJNA19867) used as reference for transcriptome assembly were downloaded from the NCBI database (www.ncbi.nlm.nig.gov).

Transcriptome assembly and mapping

After sequencing was completed, the data were transferred to a local high-performance computing server at the Center for Molecular Biology and Genetic Engineering (CBMEG, University of Campinas, Campinas, Brazil). FastQC v0.11.5 [69] was used to visually assess the quality of the sequencing reads. Removal of the remaining adapter sequences and quality trimming with a sliding window of size 4, minimum quality of 15, and length filtering (minimal length of 36 bp) was performed with Trimmomatic v0.36 [70].

The RNA-Seq data were analyzed using CLC Genomics Workbench software (v6.5.2; CLC bio, Finlandsgade, Denmark) [71]. The reads were mapped against the reference genomes of *T. harzianum* T6776 [33] and *T. atroviride* IMI 206040 [34] using the following parameters: minimum length fraction = 0.5; minimum similarity fraction = 0.8; and maximum number of hits for a read = 10. For the paired settings, the parameters were minimum distance = 150 and maximum distance = 300, including the broken pairs counting scheme. The gene expression values were expressed in reads per kilobase of exon model per million mapped reads (RPKM), and the normalized value for each sample was calculated in

transcripts per million (TPM) [72]. To statistically analyze the differentially expressed genes (DEGs), the following parameters were used: fold change greater than or equal to 1.5 or lower than or equal to -1.5 and p-value lower than 0.05.

Gene comparisons between species

Venn diagrams were constructed to compare the genes with TPM expression values greater than zero under both conditions from all species using http://bioinformatics.psb.ugent.be/webtools/Venn/.

Transcriptome annotation and CAZyme determination

Sequences were functionally annotated according to the Gene Ontology (GO) terms [73] with Blast2Go v4.1.9 [74] using BLASTx-fast and a cutoff e-value of 10^{-6}. Information derived from the CAZy database [24] was downloaded (www.cazy.org) to locally build a CAZy database (2017). The protein sequences of *T. harzianum* T6776 and *T. atroviride* IMI 206040 were used as queries in basic local alignment search tool (BLAST) searches against the locally built CAZy BLAST database. BLAST matches showing an e-value less than 10^{-11}, identity greater than 30% and queries covering greater than 70% of the sequence length were selected and classified according to the CAZyme catalytic group as GHs, CBMs, GTs, CEs, AAs or PLs and their respective CAZyme families.

CAZymes were also annotated according to Enzyme Commission (EC) number [75] through BRENDA (Braunschweig Enzyme Database) [76] (www.brenda-enzymes.org), using BLASTp with an e-value cutoff of 10^{-10}, identity greater than 30% and queries covering greater than 60% of the sequence length.

Coexpression networks

The coexpression networks for Th0179 were assembled from the mapped RNA-Seq data against the genome of *T. harzianum* T6776 using biological triplicates. The network was assembled by calculating Pearson's correlation coefficient for each pair of genes. Genes showing null values for most of the replicates under different experimental conditions were excluded to decrease noise and to remove residuals from the analysis. The highest reciprocal rank (HRR) method proposed by Mutwil et al. [77] was used to empirically filter the edges, retaining edges with an HRR less than or equal to 30. Thus, only edges representing the strongest correlations were selected. Cytoscape software v3.6.0 [78] was used for data analysis and network construction. The cluster analysis procedure was performed with the Heuristic Cluster Chiseling Algorithm (HCCA) [77].

Exoproteome analysis

The analysis of the exoproteome of Th0179 under both fermentative conditions was performed via liquid chromatography tandem mass spectrometry (LC-MS/MS) using the data-independent method of acquisition MS^E as described by Horta et al. [10] The LC-MS/MS data were processed using ProteinLynx Global Server (PLGS) v3.0.1 software (Waters, Milford, MA, USA), and the proteins in the processed files were identified by comparison to the *Trichoderma* sequence database available in the UniProt Knowledgebase (UniProtKB; https://www.uniprot.org/uniprot/). BLASTp searches of the fasta sequences of the identified proteins were performed against the *T. harzianum* T6776 genome to identify the secreted proteins and compare them to the transcriptome.

RT-qPCR analysis

To verify the reliability and accuracy of the transcriptome data and validate the differential expression results, reverse transcription-quantitative PCR (RT-qPCR) was

performed for selected genes (Additional file 8: Table S5). The RNeasy Mini Kit (Qiagen) was used for RNA extraction, and cDNA was synthesized using the QuantiTect Reverse Transcription Kit (Qiagen, Germany) according to the manufacturer's instructions. Primers were synthesized using the Primer3Plus web interface [79] with a fusion temperature between 58 and 60 °C and amplicon sizes between 120 and 200 bp.

Quantification of gene expression was performed by continuously monitoring SYBR Green fluorescence. The reactions were performed in triplicate in a total volume of 6.22 μL. Each reaction contained 3.12 μL of SYBR Green Supermix (Bio-Rad, USA), 1.0 μL of forward and reverse primers and 2.1 μL of diluted cDNA. The reactions were assembled in 384-well plates. PCR amplification-based expression profiling of the selected genes was performed using specific endogenous controls for each strain, which are described in Additional file 8: Table S5. RT-qPCR was conducted with the CFX384 Touch Real-Time PCR Detection System (Bio-Rad). The real-time PCR program was as follows: initial denaturation at 95 °C for 10 min, followed by 40 cycles of 15 sec at 95 °C and 60 sec at 60 °C. Gene expression was calculated via the delta-delta cycle threshold method [80]. The obtained RT-qPCR results were compared with the RNA-Seq results from the generated assemblies. The selected genes exhibited the same expression profiles between the RT-qPCR and RNA-Seq analyses (Additional file 3: Fig. S2).

Additional files

Additional file 1: Fig. S1. Venn diagrams. Venn diagrams of the genes identified in *Trichoderma* spp. with expression levels higher than zero under cellulose (a) and glucose (b) growth conditions using the *T. harzianum* T6776 genome as a reference.

Additional file 2: Table S1. Upregulated genes identified in *Trichoderma* spp. under cellulose growth conditions according to statistical parameters and expression levels.

Additional file 3: Fig. S2. RNA-Seq analysis validation. The obtained RT-qPCR results were compared with the RNA-Seq results for transcriptome analysis validation of *Trichoderma* spp.

Additional file 4: Table S2. Classification of the CAZyme genes under cellulose conditions for *T. harzianum* IOC-3844 and *T. atroviride* CBMAI-0020.

Additional file 5: Table S3. Proteins identified in both the transcriptome and exoproteome of *T. harzianum* CBMAI-0179 under both glucose and cellulose growth conditions.

Additional file 6: Table S4. Genes identified in the cluster analysis from the main coexpression network of *T. harzianum* CBMAI-0179.

Additional file 7: Fig. S3. Enzymatic activity. Enzymatic activities (UI mL^{-1}) of β-glucosidase (a), cellulase (b), and xylanase (c) and protein contents (d) in the culture supernatants of *Trichoderma* spp. measured after 96 h of growth. Each bar represents the mean and standard deviation of biological triplicates.

Additional file 8: Table S5. Primer sequences and amplicons of the endogenous genes evaluated in this study, and DEGs obtained in RNA-Seq data for transcriptome analysis validation by RT-qPCR.

Abbreviations

AA: Auxiliary enzymes; **BLAST:** Basic Local Alignment Search Tool; **bp:** Base pair; **BRENDA:** Braunschweig Enzyme Database; **CAZymes:** Carbohydrate-active enzymes; **CBM:** Carbohydrate-binding module; **cDNA:** Complementary DNA; **CE:** Carbohydrate esterases; **CEL:** Cellulose; **DEGs:** Differentially expressed genes; **EC:** Enzyme commission number; **FPA:** Filter paper activity; **GH:** Glycoside hydrolases; **GLU:** Glucose; **GO:** Gene Ontology; **GT:** Glycosyltransferases; **HCCA:** Heuristic Cluster Chiseling Algorithm; **HRR:** Highest reciprocal rank; **kb:** Kilobases; **LPMO:** Lytic polysaccharides monooxygenase;

MA2: Malt extract agar, 2% w/w; **Mb:** Megabase; **mL:** Milliliter; **mRNA:** Messenger RNA; **PCA:** Principal component analysis; **PDA:** Potato dextrose agar; **PL:** Polysaccharide lyases; **RNA:** Ribonucleic acid; **RNA-Seq:** RNA sequencing; **RPKM:** Reads per kilobase of exon model per million mapped reads; **RT-qPCR:** Real-time quantitative PCR; **Ta0020:** *Trichoderma atroviride* CBMAI-0020; **TFs**: Transcription factors; **Th0179:** *Trichoderma harzianum* CBMAI-0179; **Th3844:** *Trichoderma harzianum* IOC-3844; **TPM:** Transcripts per million; **UI:** International Unit; **µl:** Microliter

Acknowledgements

We would like to acknowledge the funding from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2015/09202-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Computational Biology Program) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We thank the National Institute of Metrology, Quality and Technology (INMETRO) for performing the proteomics analysis via LC-MS/MS, the Brazilian Biorenewables National Laboratory (LNBR), Campinas – SP, for conducting the fermentation experiments and the Center of Molecular Biology and Genetic Engineering (CBMEG) at the University of Campinas, SP, for use of the center and laboratory space. This manuscript was previously posted to bioRxiv https://www.biorxiv.org/content/10.1101/2020.01.14.906529v1

Authors' contributions

DAA and APS conceived and designed the study. DAA, MACH, JAFF and NFM performed the data analysis. DAA drafted the manuscript, which was critically revised by MACH, JAFF and APS. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2015/09202-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Computational Biology Program) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). DAA received an MS fellowship from FAPESP (2017/17782-2) and CAPES – Computational Biology Program (88882.160100/2017-01, 88887.336686/2019-00). MACH received a PD fellowship from FAPESP (2018/18856-1). JAFF received a PhD fellowship from CNPq (170565/2017-3). NFM received a PD fellowship from CNPq and CAPES, Computational Biology Program, and APS is the recipient of a research fellowship from CNPq. The funding bodies played no role in the design of the study, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets generated and/or analyzed during the current study are included in this published article and its Additional files 1, 2, 3, 4, 5, 6, 7 and 8. The reads have been deposited at the NCBI Sequence Read Archive (SRA) and can be accessed under the BioProject number PRJNA336221.

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

[1]Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil. [2]Graduate Program in Genetics and Molecular Biology, Institute of Biology, UNICAMP, Campinas, SP, Brazil. [3]Department of Plant Biology, Institute of Biology, UNICAMP, Campinas, SP, Brazil.

References

1. de Gouvêa PF, Bernardi AV, Gerolamo LE, Santos EDS, Riaño-Pachón DM, Uyemura SA, et al. Transcriptome and secretome analysis of *Aspergillus fumigatus* in the presence of sugarcane bagasse. BMC Genomics. 2018;19:232.
2. Castro LDS, Pedersoli WR, Antoniêto ACC, Steindorff AS, Silva-Rocha R, Martinez-Rossi NM, et al. Comparative metabolism of cellulose, sophorose and glucose in *Trichoderma* reeseiusing high-throughput genomic and proteomic analyses. Biotechnol Biofuels. 2014;7:41.
3. Santos CA, Morais MAB, Terrett OM, Lyczakowski JJ, Zanphorlin LM, Ferreira-Filho JA, et al. An engineered GH1 β-glucosidase displays enhanced glucose tolerance and increased sugar release from lignocellulosic materials. Sci Rep. 2019;9:4903.
4. 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;30:1458-80.

5. Wang Y, Fan C, Hu H, Li Y, Sun D, Wang Y, et al. Genetic modification of plant cell walls to enhance biomass yield and biofuel production in bioenergy crops. Biotechnol Adv. 2016;34:997-1017.

6. Rocha VAL, Maeda RN, Pereira N, Kern MF, Elias L, Simister R, et al. Characterization of the cellulolytic secretome of *Trichoderma harzianum* during growth on sugarcane bagasse and analysis of the activity boosting effects of swollenin. Biotechnol Prog. 2016;32:327-36.

7. Ahmed S, Mustafa G, Arshad M, Rajoka MI. Fungal biomass protein production from *Trichoderma* harzianum using rice polishing. BioMed Res Int. 2017;2017:6232793.

8. de Souza WR. Microbial degradation of lignocellulosic biomass**.** In: Chandel AK, da Silva SS, editors. Sustainable degradation of lignocellulosic biomass - techniques, applications and commercialization. London, UK: IntechOpen; 2013. p. 207-46.

9. Miao Y, Liu D, Li G, Li P, Xu Y, Shen Q, et al. Genome-wide transcriptomic analysis of a superior biomass-degrading strain of *A. fumigatus* revealed active lignocellulose-degrading genes. BMC Genomics. 2015;16:459.

10. Horta MAC, Filho JAF, Murad NF, Santos EDO, dos Santos CA, Mendes JS, et al. Network of proteins, enzymes and genes linked to biomass degradation shared by *Trichoderma* species. Sci Rep. 2018;8:1341.

11. Kumar M, Ashraf S. Role of *Trichoderma* spp. as a biocontrol agent of fungal plant pathogens**.** In: Kumar V, Kumar M, Sharma S, Prasad R, editors. Probiotics and plant health. Singapore: Springer Singapore; 2017. p. 497-506.

12. Filho JAF, Horta MAC, Beloti LL, dos Santos CA, de Souza AP. Carbohydrate-active enzymes in *Trichoderma harzianum*: a bioinformatic analysis bioprospecting for key enzymes for the biofuels industry. BMC Genomics. 2017;18:779.

13. Druzhinina IS, Seidl-Seiboth V, Herrera-Estrella A, Horwitz BA, Kenerley CM, Monte E, et al. *Trichoderma*: the genomics of opportunistic success. Nat Rev Microbiol. 2011;9:749-59.

14. Mukherjee PK, Horwitz BA, Herrera-Estrella A, Schmoll M, Kenerley CM. *Trichoderma* research in the genome era. Annu Rev Phytopathol. 2013;51:105-29.

15. Ghildiyal A, Pandey A. Isolation of cold tolerant antifungal strains of *Trichoderma* sp. from glacial sites of Indian Himalayan region. Res J Microbiol. 2008;3:559-64.

16. Peterson R, Nevalainen H. *Trichoderma reesei* RUT-C30 – thirty years of strain improvement. Microbiology. 2012;158:58-68.

17. Martinez D, Berka RM, Henrissat B, Saloheimo M, Arvas M, Baker SE, et al. Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). Nat Biotechnol. 2008;26:553-60.

18. Margolles-Clark E, Ihnen M, Penttilä M. Expression patterns of ten hemicellulase genes of the filamentous fungus *Trichoderma reesei* on various carbon sources. J Biotechnol. 1997;57:167-79.

19. Druzhinina IS, Kubicek CP. Genetic engineering of *Trichoderma reesei* cellulases and their production. Microb Biotechnol. 2017;10:1485-99.

20. Benoliel B, Torres FAG, de Moraes LMP. A novel promising *Trichoderma harzianum* strain for the production of a cellulolytic complex using sugarcane bagasse in natura. SpringerPlus. 2013;2:656.

21. Delabona PDS, Farinas CS, da Silva MR, Azzoni SF, Pradella JGDC. Use of a new *Trichoderma harzianum* strain isolated from the Amazon rainforest with pretreated sugar cane bagasse for on-site cellulase production. Bioresour Technol. 2012;107:517-21.

22. Horta MAC, Vicentini R, Delabona PDS, Laborda P, Crucello A, Freitas S, et al. Transcriptome profile of *Trichoderma harzianum* IOC-3844 induced by sugarcane bagasse. PLoS One. 2014;9:e88689.

23. Delabona PDS, Cota J, Hoffmam ZB, Paixão DAA, Farinas CS, Cairo JPLF, et al. Understanding the cellulolytic system of *Trichoderma harzianum* P49P11 and enhancing saccharification of pretreated sugarcane bagasse by supplementation with pectinase and α-l-arabinofuranosidase. Bioresour Technol. 2013;131:500-7.

24. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490-5.

25. Cairo JPLF, Leonardo FC, Alvarez TM, Ribeiro DA, Büchli F, Costa-Leonardo AM, et al. Functional characterization and target discovery of glycoside hydrolases from the digestome of the lower termite *Coptotermes gestroi*. Biotechnol Biofuels. 2011;4:50.

26. Montella S, Ventorino V, Lombard V, Henrissat B, Pepe O, Faraco V. Discovery of genes coding for carbohydrate-active enzyme by metagenomic analysis of lignocellulosic biomasses. Sci Rep. 2017;7:42623.

27. Suwannarangsee S, Bunterngsook B, Arnthong J, Paemanee A, Thamchaipenet A, Eurwilaichitr L, et al. Optimisation of synergistic biomass-degrading enzyme systems for efficient rice straw hydrolysis using an experimental mixture design. Bioresour Technol. 2012;119:252-61.

28. Villares A, Moreau C, Bennati-Granier C, Garajova S, Foucat L, Falourd X, et al. Lytic polysaccharide monooxygenases disrupt the cellulose fibers structure. Sci Rep. 2017;7:40262.

29. Bennati-Granier C, Garajova S, Champion C, Grisel S, Haon M, Zhou S, et al. Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by *Podospora anserina*. Biotechnol Biofuels. 2015;8:90.

30. Bischof RH, Ramoni J, Seiboth B. Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. Microb Cell Fact. 2016;15:106.

31. Ellegren H, Galtier N. Determinants of genetic diversity. Nat Rev Genet. 2016;17:422-33.

32. Al-Sadi AM, Al-Oweisi FA, Edwards SG, Al-Nadabi H, Al-Fahdi AM. Genetic analysis reveals diversity and genetic relationship among *Trichoderma* isolates from potting media, cultivated soil and uncultivated soil. BMC Microbiology. 2015;15:147.

33. Baroncelli R, Piaggeschi G, Fiorini L, Bertolini E, Zapparata A, Pè ME, et al. Draft whole-genome sequence of the biocontrol agent *Trichoderma harzianum* T6776. Genome Announc. 2015;3:e00647-15.

34. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, et al. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. Genome Biol. 2011;12:R40.

35. Limón MC, Chacón MR, Mejías R, Delgado-Jarana J, Rincón AM, Codón AC, et al. Increased antifungal and chitinase specific activities of *Trichoderma harzianum* CECT 2413 by addition of a cellulose binding domain. Appl Microbiol Biotechnol. 2004;64:675-85.

36. Pellegrini VOA, Serpa VI, Godoy AS, Camilo CM, Bernardes A, Rezende CA, et al. Recombinant *Trichoderma harzianum* endoglucanase I (Cel7B) is a highly acidic and promiscuous carbohydrate-active enzyme. Appl Microbiol Biotechnol. 2015;99:9591-604.

37. Valadares F, Gonçalves TA, Gonçalves DSPO, Segato F, Romanel E, Milagres AMF, et al. Exploring glycoside hydrolases and accessory proteins from wood decay fungi to enhance sugarcane bagasse saccharification. Biotechnol Biofuels. 2016;9:110.

38. 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;6:41.

39. Johansen KS. Lytic polysaccharide monooxygenases: the microbial power tool for lignocellulose degradation. Trends Plant Sci. 2016;21:926-36.

40. Alfaro M, Castanera R, Lavín JL, Grigoriev IV, Oguiza JA, Ramírez L, et al. Comparative and transcriptional analysis of the predicted secretome in the lignocellulose-degrading basidiomycete fungus *Pleurotus ostreatus*. Environ Microbiol. 2016;18:4710-26.

41. Saykhedkar S, Ray A, Ayoubi-Canaan P, Hartson SD, Prade R, Mort AJ. A time course analysis of the extracellular proteome of *Aspergillus nidulans* growing on sorghum stover. Biotechnol Biofuels. 2012;5:52.

42. Borin GP, Sanchez CC, de Santana ES, Zanini GK, dos Santos RAC, Pontes ADO, et al. Comparative transcriptome analysis reveals different strategies for degradation of steam-exploded sugarcane bagasse by *Aspergillus niger* and *Trichoderma reesei*. BMC Genomics. 2017;18:501.

43. Borin GP, Sanchez CC, de Souza AP, de Santana ES, de Souza AT, Leme AFP, et al. Comparative secretome analysis of *Trichoderma reesei* and *Aspergillus niger* during growth on sugarcane biomass. PLoS One. 2015;10:e0129275.

44. Vicentini R, Bottcher A, Brito MDS, dos Santos AB, Creste S, Landell MG, et al. Large-scale transcriptome analysis of two sugarcane genotypes contrasting for lignin content. PLoS One. 2015;10:e0134909.

45. Aranda-Martinez A, Lenfant N, Escudero N, Zavala-Gonzalez EA, Henrissat B, Lopez-Llorca LV. CAZyme content of *Pochonia chlamydosporia* reflects that chitin and chitosan modification are involved in nematode parasitism. Environ Microbiol. 2016;18:4200-15.

46. Leelavathi MS, Vani L, Reena P. Antimicrobial activity of *Trichoderma harzianum* against bacteria and fungi. Int J Curr Microbiol App Sci. 2014;3:96-103.

47. Contreras-Cornejo HA, Macías-Rodríguez L, Cortés-Penagos C, López-Bucio J. *Trichoderma virens*, a plant beneficial fungus, enhances biomass production and promotes lateral root growth through an auxin-dependent mechanism in *Arabidopsis*. Plant Physiol. 2009;149:1579-92.

48. Jang S, Kwon SL, Lee H, Jang Y, Park MS, Lim YW, et al. New report of three unrecorded species in *Trichoderma harzianum* species complex in Korea. Mycobiology. 2018;46:177-84.

49. Sharma PK, Gothalwal R. *Trichoderma*: a potent fungus as biological control agent**.** In: Singh JS, Seneviratne G, editors. Agro-environmental sustainability: volume 1: managing crop health. Cham, Switzerland: Springer International Publishing; 2017. p. 113-25.

50. Xie B-B, Qin Q-L, Shi M, Chen L-L, Shu Y-L, Luo Y, et al. Comparative genomics provide insights into evolution of *Trichoderma* nutrition style. Genome Biol Evol. 2014;6:379-90.

51. Brunner K, Zeilinger S, Ciliento R, Woo SL, Lorito M, Kubicek CP, et al. Improvement of the fungal biocontrol agent *Trichoderma* atroviride to enhance both antagonism and induction of plant systemic disease resistance. Appl Environ Microbiol. 2005;71:3959.

52. Zhao Z, Liu H, Wang C, Xu J-R. Correction: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 2014;15:274.

53. Sweeney MD, Xu F. Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: recent developments. Catalysts. 2012;2:244-63.

54. Javier PFI, Óscar G, Sanz-Aparicio J, Díaz P. Xylanases: molecular properties and applications**.** In: Polaina J, MacCabe AP, editors. Industrial enzymes: structure, function and applications. Dordrecht, Netherlands: Springer Netherlands; 2007. p. 65-82.

55. Binod P, Sukumaran RK, Shirke SV, Rajput JC, Pandey A. Evaluation of fungal culture filtrate containing chitinase as a biocontrol agent against *Helicoverpa armigera*. J Appl Microbiol. 2007;103:1845-52.

56. Manika S, Saju S, Subhash C, Mukesh S, Sharma P. Comparative evaluation of cellulase activity in *Trichoderma harzianum* and *Trichoderma reesei*. Afr J Microbiol Res. 2014;8:1939-47.

57. Lopes AM, Filho EXF, Moreira LRS. An update on enzymatic cocktails for lignocellulose breakdown. J Appl Microbiol. 2018;125:632-45.

58. Zang X, Liu M, Fan Y, Xu J, Xu X, Li H. The structural and functional contributions of β-glucosidase-producing microbial communities to cellulose degradation in composting. Biotechnol Biofuels. 2018;11:51.

59. Azevedo H, Bando S, Bertonha F, Moreira-Filho CA. Redes de interação gênica e controle epigenético na transição saúde-doença. Rev Med. 2015;94:223-9.

60. Peng M, Aguilar-Pontes MV, de Vries RP, Mäkelä MR. In silico analysis of putative sugar transporter genes in *Aspergillus niger* using phylogeny and comparative transcriptomics. Front Microbiol. 2018;9:1045.

61. Liu R, Chen L, Jiang Y, Zou G, Zhou Z. A novel transcription factor specifically regulates GH11 xylanase genes in *Trichoderma reesei*. Biotechnol Biofuels. 2017;10:194.

62. Benocci T, Aguilar-Pontes MV, Zhou M, Seiboth B, de Vries RP. Regulators of plant biomass degradation in *Ascomycetous fungi*. Biotechnol Biofuels. 2017;10:152.

63. Bailey MJ, Poutanen K. Production of xylanolytic enzymes by strains of *Aspergillus*. Appl Microbiol Biotechnol. 1989;30:5-10.

64. Zhang YHP, Hong J, Ye X. Cellulase assays**.** In: Mielenz JR, editors. Biofuels: methods and protocols. Totowa, NJ: Humana Press; 2009. p. 213-31.

65. Ghose TK. Measurement of cellulase activities. Pure Appl Chem. 1987;59:257-68.

66. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72:248-54.

67. Oliveira RR, Viana AJC, Reátegui ACE, Vincentz MGA. Short communication an efficient method for simultaneous extraction of high-quality RNA and DNA from various plant tissues. Genet Mol Res. 2015;14:18828-38.

68. Illumina. TruSeq RNA, sample preparation v2 guide. San Diego, US: Illumina; 2014.

69. Andrews S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010). Accessed 25 Mar 2017.

70. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114-20.

71. CLC Genomics Workbench. Manual for CLC genomics workbench 6.5.2 Windows, Mac OS X and Linux Denmark. Aarhus, Denmark: QIAGEN (Aarhus A/S); 2016.

72. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17:13.

73. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25:25-9.

74. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674-6.

75. Yamanishi Y, Hattori M, Kotera M, Goto S, Kanehisa M. E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs. Bioinformatics. 2009;25:i179-86.

76. Schomburg I, Jeske L, Ulbrich M, Placzek S, Chang A, Schomburg D. The BRENDA enzyme information system–from a database to an expert system. J Biotechnol. 2017;261:194-206.

77. Mutwil M, Klie S, Tohge T, Giorgi FM, Wilkins O, Campbell MM, et al. PlaNet: combined sequence and expression comparisons across plant networks derived from seven species. Plant Cell. 2011;23:895-910.

78. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498-504.

79. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to primer3. Nucleic Acids Res. 2007;35:W71-4.

80. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods. 2001;25:402-8.

Figure legends

Fig. 1. The transcriptome profiles, gene expression comparison and the main CAZyme classes identified for all strains. PCA of transcriptome mapping according to species and growth conditions (CEL – cellulose and GLU – glucose) using the *T. harzianum* T6776 genome as a reference (a), number of DEGs and differentially expressed CAZyme genes in *T. harzianum* CBMAI-0179 (Th0179), *T. harzianum* IOC-3844 (Th3844), and *T. atroviride* CBMAI-0020 (Ta0020) under cellulose growth conditions (b), the differentially expressed CAZyme classes identified for each strain under cellulose growth conditions (c).

Fig. 2. Distribution of CAZyme families in *Trichoderma* spp. Classification and quantification of CAZyme families in *T. harzianum* CBMAI-0179 (a), *T. harzianum* IOC-3844 (b), and *T. atroviride* CBMAI-0020 (c) under cellulose growth conditions.

Fig. 3. Evaluation of CAZyme family expression in *Trichoderma* spp. via RNA-Seq. Quantification of the expression of the main families related to cellulose (a) and hemicellulose (b) degradation in TPM.

Fig. 4. GO terms of *T. harzianum* CBMAI-0179 under cellulose growth conditions. The genes were annotated according to the main GO terms: molecular function (a), biological process (b), and cellular component (c).

Fig. 5. Coexpression networks of *T. harzianum* CBMAI-0179. Complete coexpression network (a), the coexpression subnetwork based on the exoproteome data (b), the enriched cluster analysis of the coexpression network (c), the subnetwork of the CAZyme genes and the secreted proteins identified in the cluster analysis (d). Red squares indicate DEGs under cellulose growth conditions, blue squares indicate DEGs under glucose growth conditions, yellow triangles indicate CAZyme genes under cellulose growth conditions, light blue

triangles indicate CAZyme genes under glucose growth conditions and purple hexagons indicate the secreted proteins under cellulose growth conditions.

Tables

Table 1. Classification of the CAZyme genes under cellulose growth conditions for *T. harzianum* CBMAI-0179

Gene ID	Protein Product	Fold Change	E-value	CAZy Classification	Enzyme Activity	EC Number	Cellulose TPM	Glucose TPM
THAR02_00068	KKP07860.1	1.87	1.00E-143	GH18 CBM1	chitinase	3.2.1.14	42.13	22.14
THAR02_00890	KKP07011.1	2.10	4.00E-15	GH3	beta-glucosidase	3.2.1.21	68.64	32.01
THAR02_01434	KKP06476.1	1.75	0	GH16	endo-1,3(4)-beta-glucanase	3.2.1.6	341.65	191.83
THAR02_01911	KKP05955.1	1.64	2.00E-53	GT4	1-acyl-sn-glycerol-3-phosphate acyltransferase	-	94.76	56.86
THAR02_02132	KKP05758.1	2.89	0	GH3	beta-glucosidase	3.2.1.21	16.39	5.56
THAR02_02133	KKP05759.1	2.27	3.00E-160	CBM1	cellulase	3.2.1.4	47.86	20.67
THAR02_02134	KKP05760.1	2.60	0	AA9 CBM1	cellulase	3.2.1.4	138.62	52.35
THAR02_02251	KKP05610.1	2.21	0	GH1	beta-glucosidase	3.2.1.21	88.05	39.02
THAR02_02560	KKP05371.1	1.97	2.00E-29	GH18	uncharacterized protein	-	151.83	75.55
THAR02_02979	KKP04958.1	2.05	3.00E-114	GH45 CBM1	cellulase	3.2.1.4	52.97	25.32
THAR02_03008	KKP04907.1	2.02	7.00E-30	GH55	glucan 1,3-beta-glucosidase	3.2.1.58	28.88	14.01
THAR02_03217	KKP04674.1	1.52	8.00E-14	GH17	hypothetical protein THAR02_03217	-	71.68	46.17
THAR02_03271	KKP04658.1	2.25	5.00E-137	GH10	endo-1,4-beta-xylanase	3.2.1.8	47.58	20.75
THAR02_03302	KKP04612.1	1.85	4.00E-95	GH16	glucan endo-1,3-beta-D-glucosidase	3.2.1.39	212.11	112.64
THAR02_04021	KKP03872.1	2.11	2.00E-84	GH72	1,3-beta-glucanosyltransferase	2.4.1.-	178.64	82.96
THAR02_04405	KKP03485.1	2.09	0	GH5 CBM1	cellulase	3.2.1.4	39.92	18.77
THAR02_04414	KKP03494.1	1.52	0	GH6 CBM1	cellulose 1,4-beta-	3.2.1.91	299.66	192.97

					cellobiosidase (nonreducing end)			
THAR02_05432	KKP02477.1	2.64	0	GH1	beta-glucosidase	3.2.1.21	125.51	46.73
THAR02_05677	KKP02215.1	40.48	2.00E-127	AA7	UDP-N-acetylmuramate dehydrogenase	-	20.74	0.50
THAR02_07531	KKP00372.1	2.05	0	GH20	beta-N-acetylhexosaminidase	3.2.1.52	1925.74	920.50
THAR02_07716	KKP00192.1	1.98	4.00E-92	GH64 CBM6	glucanase B	-	271.18	134.28
THAR02_07777	KKP00125.1	1.79	0	GH18 CBM1	chitinase	3.2.1.14	108.01	59.24
THAR02_07958	KKO99924.1	2.60	3.00E-162	GH27	alpha-galactosidase	3.2.1.22	22.70	8.58
THAR02_08897	KKO99004.1	1.73	0	GH7 CBM1	cellulose 1,4-beta-cellobiosidase (reducing end)	3.2.1.176	369.92	209.24
THAR02_10108	KKO97789.1	1.51	4.00E-29	CE9	glucosamine-6-phosphate isomerase	-	3705.87	2411.20
THAR02_10110	KKO97791.1	1.64	0	CE9	N-acetylglucosamine-6-phosphate deacetylase	3.5.1.25	6362.40	3809.76
THAR02_10273	KKO97625.1	1.62	0	GH17	glucan endo-1,3-beta-D-glucosidase	3.2.1.39	95.61	57.93

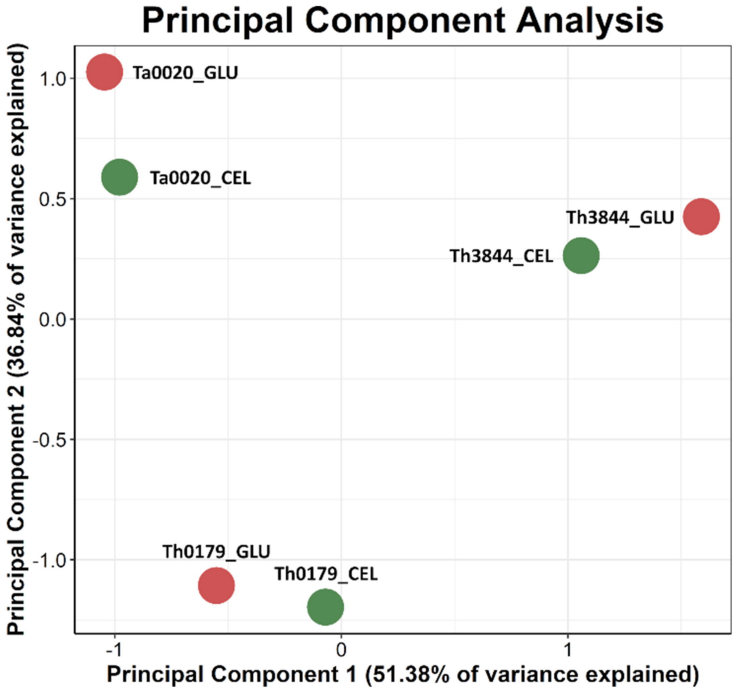
Table 2. Proteins identified in both the transcriptome and exoproteome of *T. harzianum* CBMAI-0179 grown in cellulose

Gene ID	Accession Number	Protein Name	E-value	CAZy Classification	EC Number	Cellulose TPM	Glucose TPM
THAR02_00377	G0RX84	Predicted protein	3.00E-15	-	-	1113.19	1128.71
THAR02_00656	A0A0F9XRC5	Beta-glucosidase	0	GH3	3.2.1.21	5.08	2.55
THAR02_00890	A0A0F9XQT4	Beta-glucosidase	0	GH3	3.2.1.21	68.64	32.01
THAR02_01069	G9MX73	Glycoside hydrolase family 64 protein	1.00E-64	GH64	-	17.67	25.95
THAR02_01434	A0A0F9XP75	Uncharacterized protein	0	GH16	3.2.1.6	341.65	191.83
THAR02_01982	A0A0F9Y1F6	Beta-galactosidase	0	GH35	3.2.1.23	10.08	6.85
THAR02_02133	A0A0G0AME2	Uncharacterized protein	0	CBM1	-	47.86	20.67
THAR02_02134	A0A0F9XMI8	Cellulase	0	AA9 CBM1	3.2.1.4	138.62	52.35
THAR02_02147	A0A0F9Y0Y9	Endo-1,4-beta-xylanase	0	GH11 CBM1	3.2.1.8	119.56	86.24
THAR02_02289	A0A0F9Y0G5	Cel74a	0	CBM1	-	22.53	24.27
THAR02_03210	A0A0F9ZXC9	WSC domain-containing protein	0	AA5_1	-	33.57	38.75

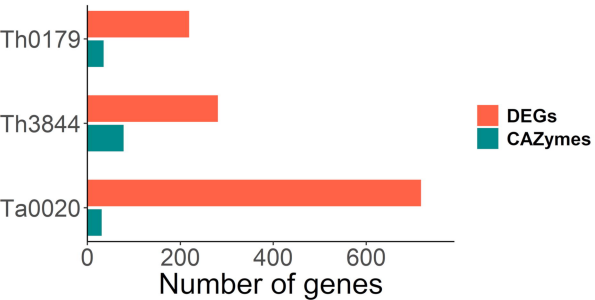
THAR02_03271	A0A0F9XXA4	Beta-xylanase	0	GH10	3.2.1.8	47.58	20.75
THAR02_03624	G0RX52	Extracellular metalloproteinase (Fungalysin)	0	CBM50	3.4.24.-	1.02	0.31
THAR02_03851	A0A0G0AGG8	Mannan endo-1,4-β-mannosidase	0	GH5 CBM1	3.2.1.78	67.98	57.54
THAR02_04062	A0A0F9XH17	Uncharacterized protein	2.00E-170	-	-	21.80	18.38
THAR02_04344	G9MY63	Glycoside hydrolase family 71 protein	0	GH71 CBM24	3.2.1.59	3.82	3.87
THAR02_04405	A0A0F9XG06	Cellulase	0	GH5 CBM1	3.2.1.4	39.92	18.77
THAR02_04414	A0A0G0AEM7	Cellulose 1,4-β-cellobiosidase (nonreducing end)	0	GH6 CBM1	3.2.1.91	299.66	192.97
THAR02_04626	G9NK86	Glycoside hydrolase family 92 protein	0	GH92	-	1.62	0.36
THAR02_04782	A0A024HVI0	Chitinase 18-5 (Fragment)	6.00E-80	GH18	-	21.24	16.93
THAR02_05380	A0A0F9XQN9	Uncharacterized protein	3.00E-171	-	-	102.99	81.78
THAR02_05501	G0R911	Glycoside hydrolase family 92	0	GH92	-	3.88	2.03
THAR02_05896	A0A0H3UCP8	Endo-1,4-beta-xylanase	8.00E-156	GH11	3.2.1.8	33.19	27.46
THAR02_06252	A0A0F9XN06	Murein transglycosylase	0	GH71 CBM24	3.2.1.59	81.33	131.74
THAR02_07321	A0A0F9X7S7	Uncharacterized protein	0	-	-	33.34	42.55
THAR02_07975	G0RXE3	Predicted protein	0	-	-	8.39	15.64
THAR02_08235	A0A0F9ZHA7	Chitinase 3	0	GH18	3.2.1.96	62.16	62.00
THAR02_08478	A0A0G0A296	Uncharacterized protein	0	GH30_7	3.2.1.-	6.37	2.67
THAR02_08479	A0A0F9X463	Uncharacterized protein	0	CBM1	-	28.08	15.80
THAR02_09247	A0A0F9ZZN6	Uncharacterized protein	0	CBM43	-	142.27	133.87
THAR02_09257	E2PTX8	Endochitinase 42 (Fragment)	6.00E-36	GH18	-	8.28	8.50
THAR02_09719	A0A0F9WYH5	Cellulase	0	GH5 CBM1	3.2.1.4	18.31	12.14

947

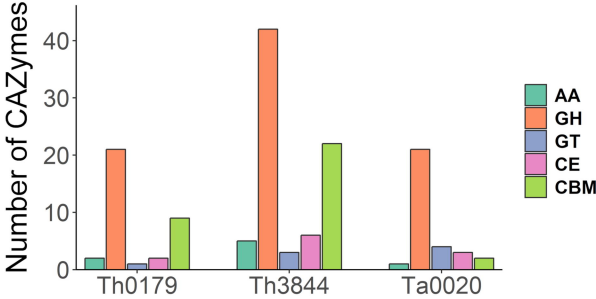
a

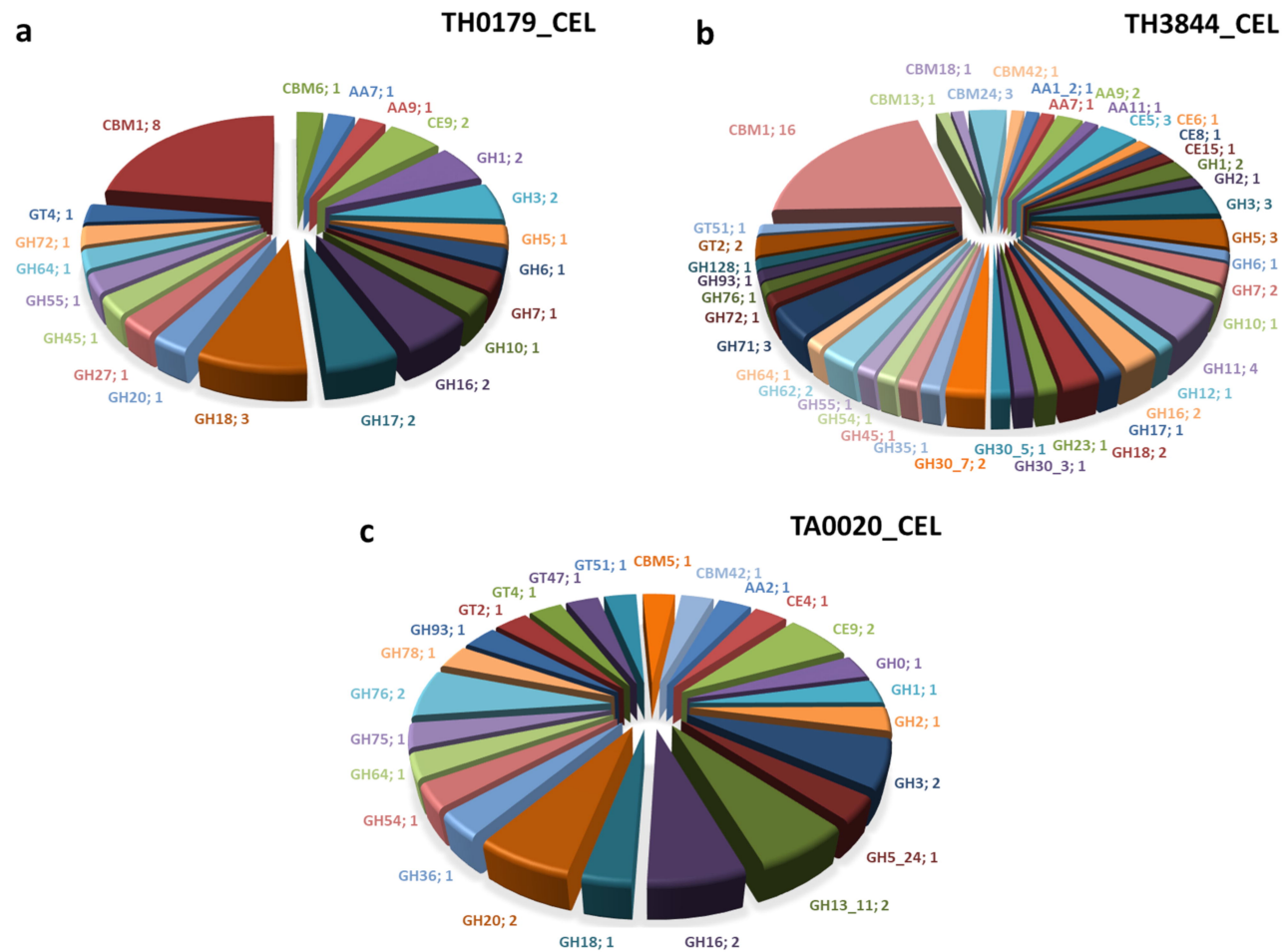

Principal Component Analysis
Ta0020_GLU
Ta0020_CEL
Th3844_GLU
Th3844_CEL
Th0179_GLU
Th0179_CEL
1.0
0.5
0.0
-0.5
-1.0
-1
0
1
Principal Component 2 (36.84% of variance explained)
Principal Component 1 (51.38% of variance explained)

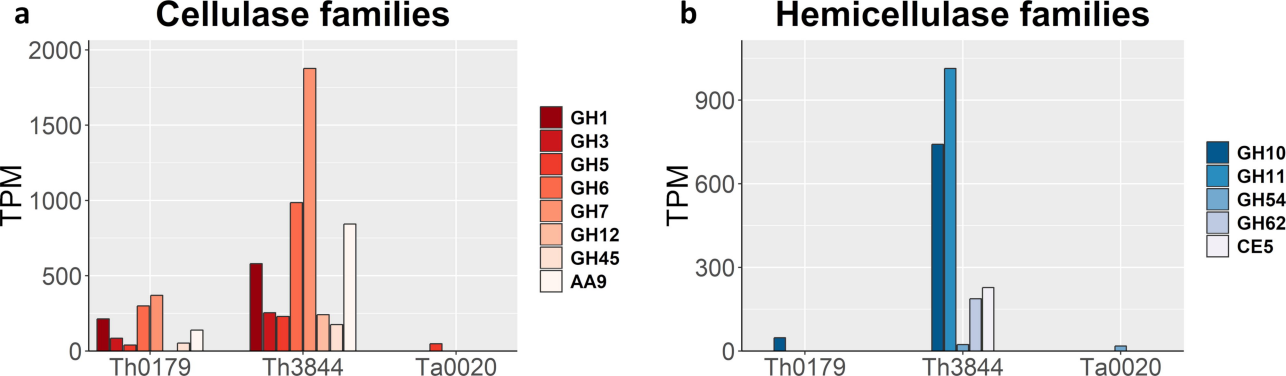
b


Th0179
Th3844
Ta0020
0
200
400
600
Number of genes
DEGs
CAZymes

c

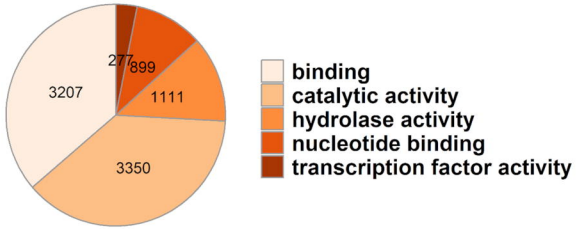

Number of CAZymes
0
10
20
30
40
Th0179
Th3844
Ta0020
AA
GH
GT
CE
CBM


a
TH0179_CEL
CBM6; 1
AA7; 1
AA9; 1
CE9; 2
GH1; 2
GH3; 2
GH5; 1
GH6; 1
GH7; 1
GH10; 1
GH16; 2
GH17; 2
GH18; 3
GH20; 1
GH27; 1
GH45; 1
GH55; 1
GH64; 1
GH72; 1
GT4; 1
CBM1; 8
b
TH3844_CEL
CBM18; 1
CBM42; 1
CBM13; 1
CBM24; 3
AA1_2; 1
AA9; 2
AA7; 1
AA11; 1
CE5; 3
CE6; 1
CE8; 1
CE15; 1
GH1; 2
GH2; 1
GH3; 3
GH5; 3
GH6; 1
GH7; 2
GH10; 1
GH11; 4
GH12; 1
GH16; 2
GH17; 1
GH18; 2
GH23; 1
GH30_3; 1
GH30_5; 1
GH30_7; 2
GH35; 1
GH45; 1
GH54; 1
GH55; 1
GH62; 2
GH64; 1
GH71; 3
GH72; 1
GH76; 1
GH93; 1
GH128; 1
GT2; 2
GT51; 1
CBM1; 16
c
TA0020_CEL
CBM5; 1
CBM42; 1
AA2; 1
CE4; 1
CE9; 2
GH0; 1
GH1; 1
GH2; 1
GH3; 2
GH5_24; 1
GH13_11; 2
GH16; 2
GH18; 1
GH20; 2
GH36; 1
GH54; 1
GH64; 1
GH75; 1
GH76; 2
GH78; 1
GH93; 1
GT2; 1
GT4; 1
GT47; 1
GT51; 1


a
Cellulase families
TPM
2000
1500
1000
500
0
Th0179
Th3844
Ta0020
GH1
GH3
GH5
GH6
GH7
GH12
GH45
AA9
b
Hemicellulase families
TPM
900
600
300
0
Th0179
Th3844
Ta0020
GH10
GH11
GH54
GH62
CE5

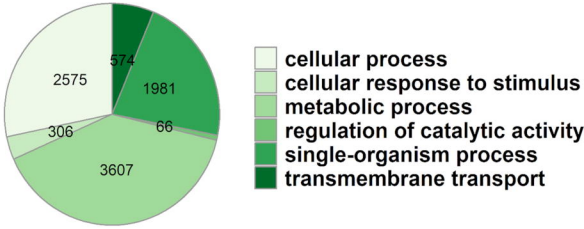
Molecular Function

a



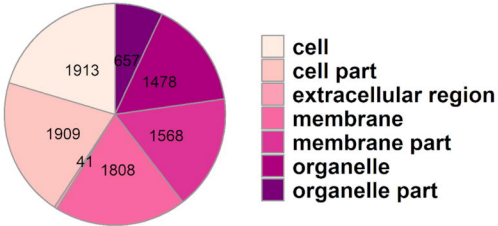
Biological Process

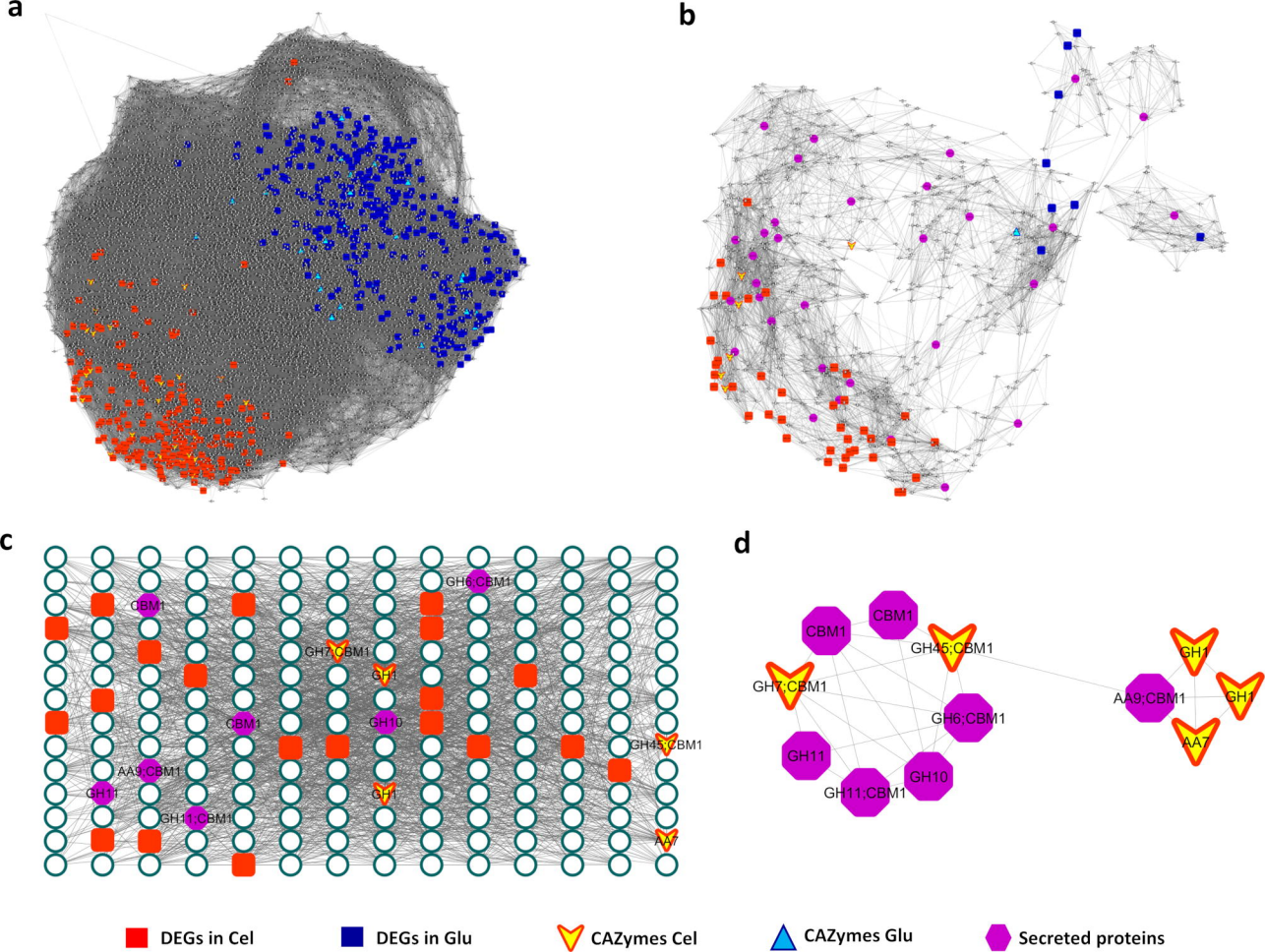
b



Cellular Component

c




a
b
c
d
CBM1
GH6;CBM1
GH7;CBM1
GH1
GH10
GH45;CBM1
AA9;CBM1
GH11
GH11;CBM1
AA7
GH45;CBM1
GH7;CBM1
GH6;CBM1
GH11
GH10
GH11;CBM1
AA9;CBM1
DEGs in Cel
DEGs in Glu
CAZymes Cel
CAZymes Glu
Secreted proteins