Comparative gene coexpression networks analysis reveals different strategies of *Trichoderma* spp. associated with XYR1 and CRE1 during cellulose degradation

Rafaela Rossi Rosolen[1,2], Alexandre Hild Aono[1,2], Déborah Aires Almeida[1,2], Jaire Alves Ferreira Filho[1,2], Maria Augusta Crivelente Horta[1,2] 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

***Corresponding author**

Prof 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: *Trichoderma* species have attracted attention as alternative plant polysaccharide-degrading enzyme sources, among other reasons. Most *Trichoderma* spp., such as *Trichoderma atroviride* and *Trichoderma harzianum*, are described as mycoparasitic fungi and are widely used in agriculture as biocontrol agents. *T. harzianum* is also a potential candidate for hydrolytic enzyme production in which gene expression is tightly controlled. Thus, a better understanding of transcriptional regulation related to the main activator (XYR1) and repressor (CRE1) of cellulases and hemicellulases is necessary. Herein, we explore the genetic mechanisms of the key regulators of genes encoding plant cell wall-degrading enzymes by inferring a gene coexpression network for *T. harzianum* (IOC-3844 and CBMAI-0179) and, for comparative purposes, for *T. atroviride* CBMAI-0020.

Results: Phylogenetic analysis indicated that both regulatory proteins are widely distributed among ascomycete fungi and suggested how *T. atroviride* is differentiated from *T. harzianum*. The coexpression network analyses separated the transcripts into several groups according to the expression profile during growth in cellulose or glucose. Groups with *xyr1* or *cre1* were identified, and within them, transcripts encoding carbohydrate-active enzymes (CAZymes), TFs, sugar and ion transporters, and proteins with unknown function were coexpressed. However, qualitative and quantitative differences among groups and strains were observed. Core transcripts from these groups (hubs) were identified, and they included transcripts not yet characterized or described as related to cellulose degradation. Several metabolic pathways triggered were recognized with high similarity between both regulators during cellulose degradation, but they differed according to the strains. An enrichment related to mycoparasitism in *T. atroviride*, especially that associated with CRE1, was observed. It was also noticed that the transcripts encoding CAZymes related to polysaccharide degradation in *T. harzianum* are not necessarily coexpressed with the TFs studied herein.

Conclusion: Our results suggest that different strategies related to XYR1 and CRE1 are used by *Trichoderma* spp. during cellulose degradation and provide new insights into transcripts coexpressed with both TFs in *T. harzianum*. This knowledge can be exploited to improve the understanding of the genetic mechanisms involved in hydrolytic enzyme regulation, expanding the potential of *T. harzianum* use in several industrial applications.

Keywords: *Trichoderma harzianum*, *Trichoderma atroviride*, cellulose, XYR1, CRE1, weighted gene coexpression network analysis.

Background

Due to the expanding population and industrialization, the demand for energy is increasing throughout the world. Currently, the primary energy source is fossil fuels, i.e., nonrenewable sources such as natural gas, oil, and coal [1]. To reduce the dependency on these finite resources, the development of new sustainable alternatives has become critical. In this context, increasing attention has been given to the development of biofuels from biomass [2, 3].

The United States and Brazil are the world's largest bioethanol producers and are responsible for 56 and 28% of its global production, respectively [4]. In Brazil, the sugar and first-generation biofuel (1G ethanol) production system uses sugarcane as feedstock, which generates large amounts of lignocellulosic materials as residue [5]. On average, 280 kg of bagasse is produced through industrial processing of 1 ton of sugarcane [6]. Thus, new technologies based on converting agricultural residues into second-generation biofuel (2G ethanol) are emerging [7-9].

Despite advances, the cost of biomass-degrading enzymes remains the main bottleneck for commercial biofuel production [10-12]. In nature, bacteria and fungi produce and secrete a combination of enzymes that act synergistically to break down complex polysaccharides [13]. To establish 2G ethanol as a competitive technology, the application of these enzymes for the improvement of enzymatic cocktails is earning great attention in the biofuel industry [14, 15].

Because of the exclusive nutritional versatility, species of the filamentous ascomycete genus *Trichoderma* have been used in a wide range of biotechnological applications [16-19]. While fungal cellulases from *Trichoderma reesei* are the most applied enzymes in the biotechnology industry [17, 20], the use of *Trichoderma atroviride* and *Trichoderma virens* has been explored by examining their biocontrol capacity against plant pathogenic fungi [21, 22]. Both species are distantly related to *T. reesei* [23] and represent well-defined phylogenetic species [24, 25]. In contrast, *Trichoderma harzianum sensu lato* is also commonly used as a biocontrol agent but is part of a complex of several cryptic species [26]. More recently, studies on *T. harzianum* strains have shown their potential to produce a set of enzymes that can degrade the residual biomass of sugarcane [27-32].

Lignocellulosic biomass is a complex recalcitrant structure that requires a consortium of enzymes for its complete depolymerization. For cellulose degradation, three types of cellulases are required: cellobiohydrolases, endoglucanases and β-glucosidases. For hemicellulose hydrolysis, several enzymes are needed, such as endo-xylanases, β-xylosidases, β-mannanases, arabinofuranosidases, α-glucuronidases and esterases [33]. Furthermore, auxiliary enzymes and proteins are involved in the degradation of the plant wall, such as lytic polysaccharide monooxygenases (LPMOs) [34] and swollenin [35], which can improve the overall performance of biomass hydrolysis [29].

These hydrolytic enzymes and other accessory proteins related to complete breakdown of cellulose and hemicellulose are collectively named carbohydrate-active enzymes (CAZymes) [36]. The expression of CAZyme-encoding genes is regulated at the transcriptional level in a coordinated manner that depends on the availability of the carbon sources [37-39]. In *T. reesei*, cellulase and hemicellulase gene expression is regulated by the action of several transcription factors (TFs) [40]. The main TF involved in the expression of most cellulase and hemicellulase genes is xylanase regulator 1 (XYR1) [40, 41]. Moreover, the expression of these genes is subject to carbon catabolite repression (CCR) [42], which is regulated by carbon catabolite repressor 1 (CRE1) [43].

Although XYR1 orthologs are present in almost all filamentous ascomycete fungi, the molecular mechanisms triggered by this regulator depend on the species [44]. XYR1 was first described in *T. reesei* as the main activator of enzymes involved in d-xylose metabolism [41]. However, for *T. atroviride*, the induction of genes encoding cell wall-degrading enzymes considered relevant for mycoparasitism, such as *axe1* and *swo1*, is influenced by XYR1 [45]. CRE1 is the only conserved TF throughout the fungal kingdom, suggesting a conserved mechanism for CCR in fungi [46], and in many cases, it has a negative effect on the production of lignocellulose-degrading enzymes [47].

One of the great challenges in understanding the molecular mechanisms of expression of CAZymes and other proteins required for biomass deconstruction is how the main TFs act. Given that inside the genus *Trichoderma*, research on the gene regulation of lignocellulose degrading enzymes has focused on *T. reesei* [48-50], a better understanding of the transcriptional regulation of another *Trichoderma* spp. has been shown to be necessary to optimize the use of these fungi as enzyme producers.

One interesting approach to explore the regulation of lignocellulolytic enzymes *in silico* is through gene coexpression networks, which have proven to be a fundamental tool to indicate the transcripts that are active simultaneously; thus, they may have a common function [51, 52].

To date, only a few studies have employed the gene coexpression network in studies on *Trichoderma* spp. [53-55]. Through this methodology, no research has comparatively explored the genetic mechanisms triggered by the main regulators involved in plant biomass degradation in *Trichoderma* spp. In this context, we considered a common mycoparasitic ancestor species (*T. atroviride* CBMAI-0020) and two mycoparasitic strains with hydrolytic potential (*T. harzianum* CBMAI-0179 and *T. harzianum* IOC-3844) [54, 55] to model a network for each strain. This approach allowed us to identify and compare modules, hub genes and metabolic pathways associated with CRE1 and XYR1 under cellulose degradation. A better understanding of the regulation of cell wall-degrading enzymes will help to elucidate the events that trigger cellulose degradation in mycoparasite *Trichoderma* species. This will be important to increase the efficiency of the production of these enzymes and expand the potential of *T. harzianum* in several industrial applications, including in the biofuel industry.

Materials and Methods

Fungal Strains and Culture Conditions

The species originated from the Brazilian Collection of Environment and Industry Microorganisms (CBMAI), located in the Chemical, Biological and Agricultural Pluridisciplinary Research Center (CPQBA) at the University of Campinas (UNICAMP), Brazil. Briefly, *T. harzianum* CBMAI-0179 (Th0179), *T. harzianum* IOC-3844 (Th3844) and *T. atroviride* CBMAI-0020 (Ta0020) strains were cultivated on solid medium for 8 days at 37 °C as described in a previous work and scraped with a washing solution [54]. The fermentation process was performed in biological triplicates and was initiated with the inoculation of 10^7 spores/mL per 200 mL of preinoculum composed of 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 [54]. Glucose was used as a control in RNA Sequencing (RNA-Seq) experiments. After 72 h of incubation, 50 mL of the preinoculum was used to inoculate 450 mL of fermentation solution in a 2 L Erlenmeyer flask. The

fermentation process continued for 96 h. Then, mycelial samples from cellulose and glucose conditions were extracted from the Th0179, Th3844 and Ta0020 strains, stored at -80 °C and ground in liquid nitrogen. Frozen material was used for RNA extraction, and the sequencing experiment was carried out on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA). These data were obtained and described in a previous study [54].

Transcription Profiling

Sequencing quality filters were applied using Trimmomatic v0.36 [56] as described by [55]. Reads were mapped against *T. harzianum* T6776 [57] and *T. atroviride* IMI206040 [58] reference genomes using CLC Genomics Workbench software (CLC bio – 6.5.2 v; Finlandsgade, Dk) [59]. Gene expression values were calculated in transcripts per million (TPM), and genes with p-values ≤ 0.05 and fold changes ≥ 1.5 (upregulated) or ≤ -1.5 (downregulated) were considered differentially expressed, as in [55].

Phylogenetic Analysis of CRE1 and XYR1 Proteins

The ascomycete fungus CRE1 and XYR1 protein sequences with similarity ≥ 60% and ≥ 70%, respectively, to the reference genome of *T. harzianum* T6776 were retrieved from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/). Additionally, from the selected genome, CRE1 and XYR1 sequences were recovered and used as a reference for Th3844, Th0179 and Ta0020 TF consensus sequences in RNA-Seq read mapping. These sequences were also included in the phylogenetic analyses. Multiple sequence alignment was performed using ClustalW [60], and a phylogenetic tree was created using Molecular Evolutionary Genetics Analysis (MEGA) software v7.0 [61]. The maximum likelihood (ML) [62] method of inference was used based on the Dayhoff model with a freqs (F^+) matrix-based model for CRE1 and based on the Jones-Taylor-Thornton (JTT) model with freqs (F^+) for XYR1. We used 1,000 bootstrap replicates [63] for each analysis. Pairwise deletion was employed to address alignment gaps and missing data. Trees were visualized and edited using the FigTree program v1.4.3 (http:/tree.bio.ed.ac.uk/software/figtree/).

Gene Coexpression Network Construction

The gene coexpression networks were modeled for Th3844, Th0179 and Ta0020 using transcriptome data (PRJNA336221) from a previous work that were obtained with biological triplicates using cellulose or glucose as the carbon source [54]. For the network construction, TPM values were used along with the R package for weighted correlation network analysis (WGCNA) [64]. To decrease noise and to remove residuals from the analysis, WGCNA filters were applied to remove transcripts with zero variance for any replicate under different experimental conditions and a minimum relative weight of 0.1. Pearson's correlation values among all pairs of transcripts across all selected samples for each strain were used for network construction. A softpower β was chosen for each network using *pickSoftThreshold* to fit the signed network to a scale-free topology to obtain an adjacency matrix where nodes correspond to genes and edges to the strength of their connection. To obtain a dissimilarity matrix, we built a topological overlap matrix as implemented in the package, with the exponentiation performed on each edge value.

Identification of Groups, Subgroups and Hubs

To identify groups of transcripts densely connected in the gene coexpression network, we used simple hierarchical clustering on the dissimilarity matrix with R [65]. From the acquired dendrogram, the dynamicTreeCut package [66] was used to obtain the ideal number of groups and the respective categorization of the transcripts. According to the functional annotation performed by [55], groups containing the desired TFs XYR1 and CRE1 were identified and named the *xyr1* and *cre1* groups, respectively. Within each detected group associated with different TFs, we used Python 3 programming language [67] together with the networkx library [68] to identify subgroups using the Louvain method [69] and hubs based on the degree distribution. We used different thresholds for keeping edges on the networks and recognizing hubs: for the *cre1* group, (I) 0.9 (all strains) was used, and for the *xyr1* group, (II) 0.9 (Th3844), (III) 0.5 (Th0179) and (IV) 0.7 (Ta0020) were used. Additionally, we tested the Molecular Complex Detection (MCODE) (v1.4.2) plugin [70] using Cytoscape software v3.7.2 with the parameters set to default (degree cutoff = 2, node score cutoff = 0.2, K-core = 2, maximum depth = 100, haircut method) for subgroup identification.

Functional Annotation of Transcripts within *xyr1* and *cre1* Groups

Functional annotation of the *xyr1* and *cre1* groups was performed based on an analysis of their transcript composition. Gene Ontology (GO) [71] terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) [72] Orthologies (KOs) related to the selected groups were chosen to explore the biological functions and metabolic pathways in which these transcripts participate. All identified GO categories were used to create a treemap to visualize possible correlated categories in the dataset caused by cellulose degradation. This step was performed using the REVIGO tool [73]. The metabolic pathways related to the KOs obtained and found in *T. reesei* were then selected due the high number of annotations correlated to this species in the KEGG database. To obtain the pathways related to the enzymes identified as belonging to the different groups, we used the Python 3 programming language [67] together with the BioPython library [74]. We selected metabolic pathways linked to the TFs XYR1 and CRE1 that act in categories related to biomass degradation in *Trichoderma* species.

Results

Molecular Phylogeny of CRE1 and XYR1 in Ascomycota

The phylogenetic analysis of the amino acid sequences related to CRE1 and XYR1 showed a clear diversification of these regulators inside the ascomycete fungi (**Fig. 1** and Additional file 1: Table S1). The genus *Trichoderma* is a member of the family Hypocreaceae, and we observed in both phylogenetic trees the separation of this family from the other families classified in the phylum Ascomycota (with bootstrap support of 100% in both cases). However, the families observed in the CRE1 phylogenetic tree (**Fig. 1a**) were not the same as those identified in XYR1 (**Fig. 1b**) as well the phylogenetic relationships of the complete protein sequences related to both TFs of the studied *Trichoderma* species/strains.

The CRE1 phylogenetic tree showed a high phylogenetic proximity between Th3844 and Th0179 strains (with bootstrap support of 51%). For both *T. harzianum* strains, the protein sequence related to CRE1 presented a close relationship with the corresponding sequence in other species of the same genus, such as *T. virens* (XP_013956509.1), *Trichoderma guizhouense* (OPB38360.1) and *Trichoderma lixii* (CAA64656.1), as well in other *T. harzianum* (PNP52054.1, KKO97123.1 and XP_024780493.1) (with bootstrap support of 95%) (**Fig. 1a**). On the other hand, Ta0020 shared closer affinity with two other *T. atroviride* strains (XP_013941427.1 and AKM12665.1), with 99% bootstrap support (**Fig. 1a**).

We observed in the phylogenetic tree of XYR1 a high phylogenetic proximity between Th3844 and *T. harzianum* CBS 226.95 (XP_024780881.1) (with bootstrap support of 96%) (**Fig. 1b**), while Th0179 was closer to *T. harzianum* T6776 (KKO98630.1) (with bootstrap support of 50%) (**Fig. 1b**) and Ta0020 was closer to *T. atroviride* IMI206040 (XP_013941705.1) (with bootstrap support of 88%) (**Fig. 1b**).

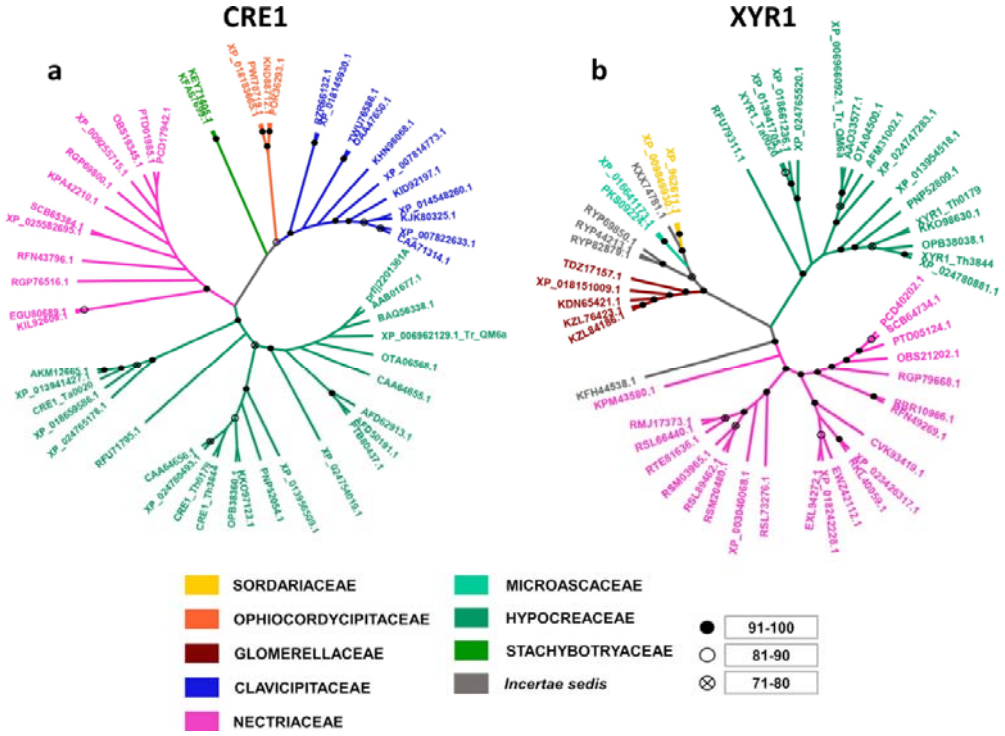

Fig. 1. Molecular phylogeny of CRE1 and XYR1 in Ascomycota. The complete protein sequences related to CRE1 (**a**) and XYR1 (**b**) were used to analyze and compare the phylogenetic relationships of the studied *Trichoderma* species/strains inside the ascomycete fungi. The amino acid sequences for Th3844, Th0179 and Ta0020 were recovered from the selected genome and used as a reference for consensus sequences in RNA-Seq read mapping. The other amino acid sequences were derived from the NCBI database. The phylogenetic trees were drawn using MEGA7 software and edited using the FigTree program.

Construction of Gene Coexpression Networks for *Trichoderma* spp.

We used transcriptome data obtained in a previous work [55] for Th3844, Th0179 and Ta0020. Through this dataset, gene coexpression networks were modeled for each strain. Transcript identification was obtained from comparative alignments of sequencing data and the following reference sequences: the *T. harzianum* T6776

genome for Th3844 and Th0179 and the *T. atroviride* IMI206040 genome for Ta0020 [55]. Using these data and applying the described filters to the WGCNA package, we obtained (I) 11,050 transcripts (Th3844), (II) 11,105 transcripts (Th0179), and (III) 11,021 transcripts (Ta0020) (Additional file 2: Table S2). Based on these transcripts, we calculated Pearson's correlation matrices. Different softpower β values were chosen to fit the scale independence of the different networks to a scale-free topology (27 for Ta0020, 8 for Th0179, and 48 for Th3844) by considering the connectivity between the transcripts based on their expression correlation. Based on transcript expression levels, the networks were partitioned into manageable groups to explore the putative coregulatory relationships. In total, we identified 87 groups for Th3844, 75 groups for Th0179 and 100 groups for Ta0020 (Additional file 3: Table S3).

Coexpressed Transcripts with XYR1 and CRE1

Our focus in *Trichoderma* spp. coexpression networks were the TFs XYR1 and CRE1, two regulators that play a central role in the expression of cellulase and hemicellulase genes during biomass degradation [46]. Therefore, transcripts encoding these proteins were sought within each module generated using the WGCNA package. Transcripts having correlated expression patterns with these regulators were identified in the Th3844, Th0179 and Ta0020 coexpression networks and grouped together (Additional file 4: Table S4). In relation to the number of transcripts, Ta0020 presented the highest coexpression with *cre1* and the lowest coexpression with *xyr1* compared to the number of transcripts identified in the other strains, which showed the opposite profile (**Table 1**). Between *T. harzianum* strains, the amount of coexpressed transcripts related to *xyr1* and *cre1* differed (**Table 1**).

Table 1. Total numbers of coexpressed transcripts identified in *Trichoderma* spp. distributed in the *cre1* and *xyr1* groups.

Groups	**Strains**		
	Th3844	**Th0179**	**Ta0020**
cre1	53	116	251
xyr1	167	153	50
Total	220	269	301

Each group previously identified containing the *xyr1* and *cre1* transcripts was partitioned, and subgroups were formed for all strains. Thus, we could identify the transcripts that are most correlated inside the modules. We tested the MCODE methodology as previously used for gene coexpression networks [53]. The number of subgroups identified was (I) 5 for Ta0020 (the subgroups contained 23, 10, 23, 16 and 6 transcripts), (II) 4 for Th0179 (the subgroups contained 3, 3, 5 and 4 transcripts), (III) 3 for Th3844 (the subgroups contained 20, 14, and 9 transcripts) for subgroups related to the *cre1* groups, (IV) 1 for Ta0020 (8 transcripts), (V) 4 for Th0179 (the subgroups contained 56, 18, 16 and 7 transcripts), and (VI) 2 for Th3844 (the subgroups contained 124 and 43 transcripts) for subgroups related to the *xyr1* groups (Additional file 5: Table S5).

We sought the transcripts encoding CRE1 and XYR1 inside these subgroups and found them for only Th3844 (*cre1* and *xyr1*) and Th0179 (*cre1*). Although we found transcripts encoding CAZymes, transporters, TFs and other interesting proteins, the MCODE methodology was not useful for all the modeled coexpression networks. Thus, we tested the Louvain algorithm [69] for detecting the most correlated niches in the *cre1* and *xyr1* groups. This approach enabled the identification of the main subgroups for the molecular reactions occurring during cellulose degradation (**Fig. 2** and Additional file 6: Table S6).

As shown in **Fig. 2**, each strain presented a different coexpression subnetwork topology for the *cre1* and *xyr1* groups based on the distribution of the subgroups. For the *cre1* groups, the Th0179 subnetwork is more tightly interconnected than the Ta0020 subnetwork and is less than the Th3844 subnetwork. For the *xyr1* groups, the same profile was observed for all strains. For each subgroup formed inside a group, a color was attributed to the transcripts that constituted the respective subgroup (Additional file 7: Table S7).

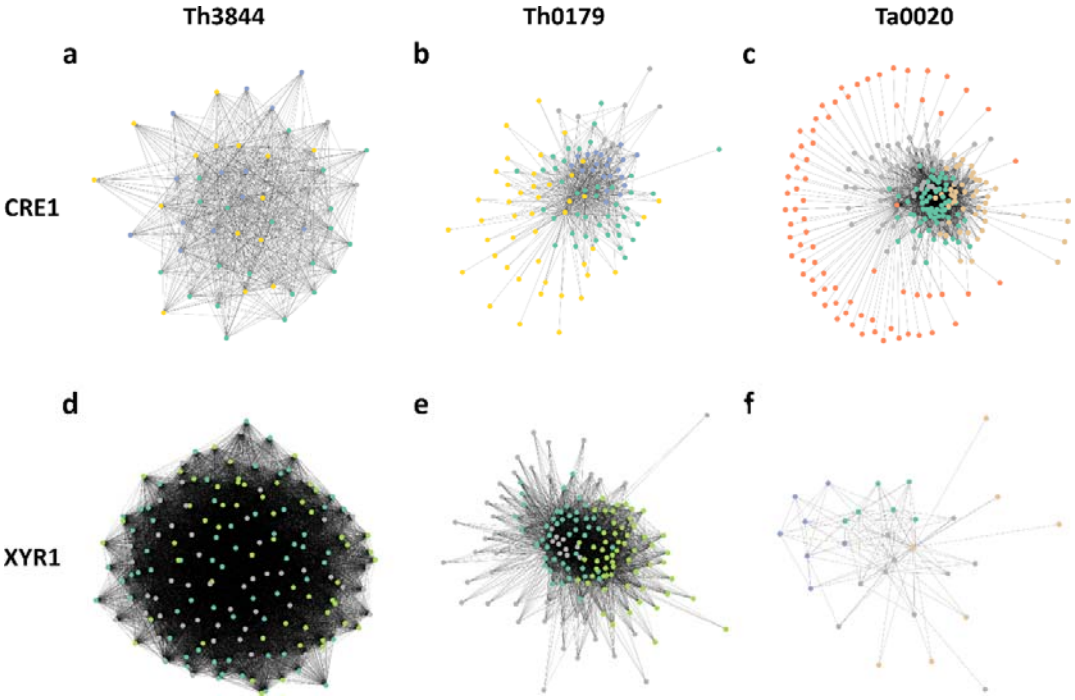

Fig. 2. Subgroups identified in the *cre1* and *xyr1* groups among *Trichoderma* spp. The groups containing CRE1 and XYR1 were partitioned according to the Louvain algorithm, and subgroups were formed for the Th3844 (**a**), Th0179 (**b**) and Ta0020 (**c**) groups related to *cre1* and for the Th3844 (**d**), Th0179 (**e**) and Ta0020 (**f**) groups related to *xyr1*. Different colors represent different subgroups inside a group. Among groups, equal colors do not represent the same transcripts. Th3844: *T. harzianum* IOC-3844; Th0179: *T. harzianum* CBMAI-0179; Ta0020: *T. atroviride* CBMAI-0020.

For each subgroup formed from the coexpressed transcripts in the *cre1* and *xyr1* groups, a denotation was attributed according to the strain (Additional file 7: Table S7). The *cre1* transcript was in the **1.0**, **2.1** and **3.5** subgroups for Th3844, Th0179 and Ta0020, respectively (Additional file 7: Table S7). We found the *xyr1* transcript associated with the transcripts that formed the **4.1**, **5.2** and **6.5** subgroups for Th3844, Th0179 and Ta0020, respectively (Additional file 7: Table S7). In addition, we observed the presence of disconnected elements in the subgroups formed for Th0179 and Ta0020 in the *cre1* group. This same pattern was observed in the *xyr1* group for Ta0020 (Additional file 7: Table S7). For the *cre1* subgroup, Ta0020 showed more than double the number of transcripts identified in the *T. harzianum* strains, and for the *xyr1* subgroup, it presented a few transcripts (Additional file 7: Table S7).

Hub Transcripts with Expression Patterns Similar to XYR1 and CRE1

The predictive potential of transcripts with high connectivity in gene coexpression networks involved in cellulose degradation by *Trichoderma* spp. remains to be thoroughly investigated. In this way, transcripts with a high number of coexpressed neighbors, defined here as hubs, were sought in groups in which the transcripts *cre1* and *xyr1* were identified (Additional file 8: Table S8). In some cases, we had to decrease the size of the correlation coefficient threshold to identify highly connected transcripts within a group. Thus, different thresholds representing cases from a moderate correlation (0.5) to a very high correlation (0.9) [75] were used. To avoid extrapolating inferences, only transcripts with a degree of connectivity were selected to improve understanding.

In the *cre1* group, the transcripts encoding a concentrative nucleoside transporter (CNT) (TRIATDRAFT_46358) for Ta0020, a mitochondrial ribosomal protein (MRP) family L3 (THAR02_01324) for Th0179 and a proteasome component PRE3 precursor (THAR02_10705) and CRE1 (THAR02_10775) for Th3844 were identified as hubs, with the last having a second candidate. In the *xyr1* group, transcripts encoding unknown protein were found to be hubs for Ta0020 (TRIATDRAFT_40550), Th0179 (THAR02_07681) and Th3844 (THAR02_00323). A transcript encoding XYR1 (TRIATDRAFT_78601) was also defined as a hub for Ta0020.

Functional Annotation of Transcripts within *xyr1* and *cre1* Groups

GO enrichment analysis was performed with the transcripts within each *cre1* and *xyr1* group. For the *cre1* group, under the molecular function category, catalytic activity (GO: 0003824) was the main annotated function for Th3844, Th0179 and Ta0020. Other functions that were enriched among these species under this category were hydrolase activity (GO: 0016787), nucleic acid binding (GO: 0003676), oxidoreductase activity (GO: 0016491) and metal ion binding (GO: 0046872). For the biological process category, the main function was metabolic process (GO: 0008152). Other functions observed in all strains that deserve special attention under this category were oxidation-reduction process (GO: 0055114) and transcription/DNA-templated (GO: 0006351). In addition, for the cellular component category, the main function was integral component of membrane (GO: 0016021), followed by cytoplasm (GO: 0005737) and nucleus (GO: 0005634).

For the *xyr1* group, catalytic activity was the main annotated function for all strains under the molecular function category. Other enriched functions found in *T. harzianum* strains were ATP binding (GO: 0005524), DNA binding (GO: 0003677), zinc ion binding (GO: 0008270) and oxidoreductase activity. It is important to note that hydrolase activity was the second main annotated function under the molecular function category for Th3844, whereas for Th0179 and Ta0020, this activity was not enriched. For the biological process category, the main function was metabolic process. Other functions shared by Th3844, Th0179 and Ta0020 enriched under this category were oxidation-reduction process, transcription/DNA-templated and regulation of transcription/DNA-templated (GO: 0006355). For the cellular component, the main function was integral component of membrane, followed by nucleus.

KO functional annotation of the transcripts encoding enzymes and unknown proteins with enzymatic activity enriched in the *cre1* and *xyr1* groups was also investigated (**Fig. 3**). For the 14 pathway classes observed for each TF, 13 were shared between the two regulators. Only two pathway classes, metabolism of other amino acids and lipid metabolism, were exclusive to the TFs CRE1 and XYR1, respectively. However, the number of identified enzymes for a given pathway class was different for each strain (**Fig. 3**).

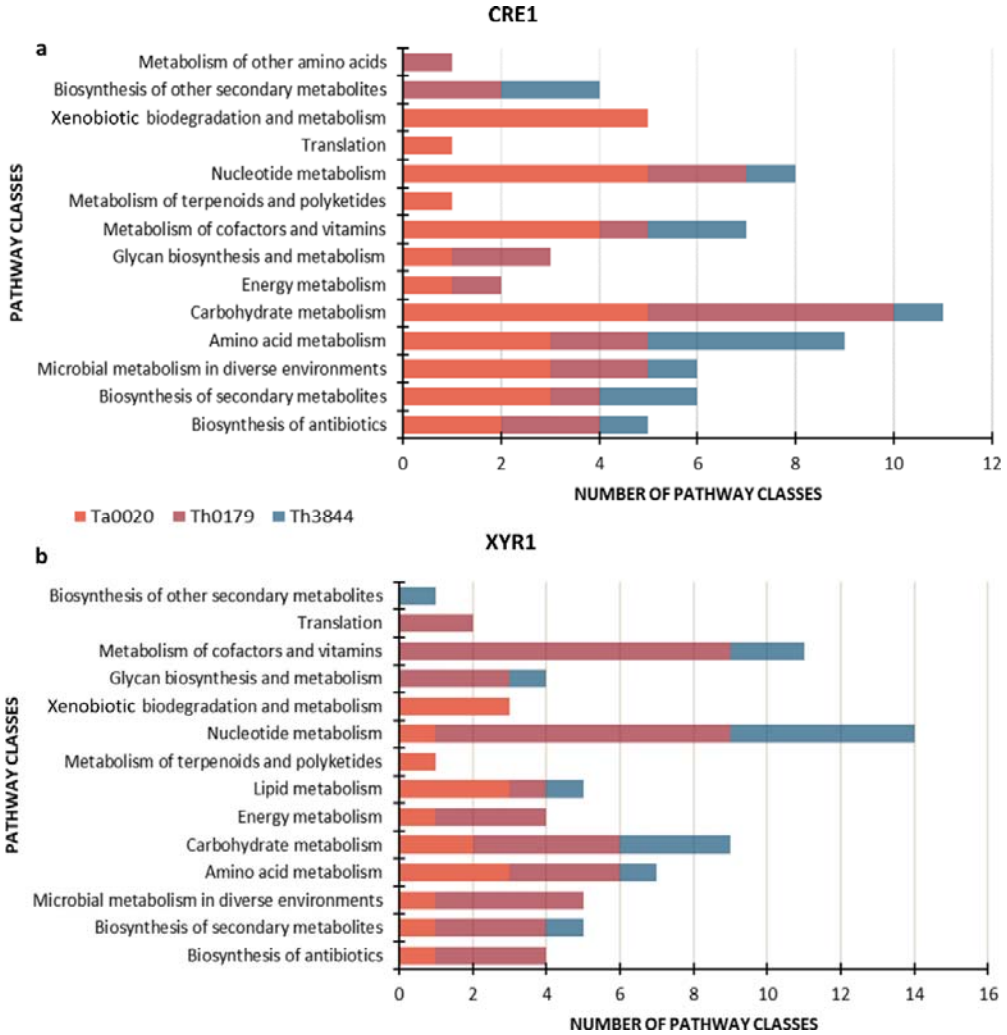

Fig. 3. KO functional classification of the transcripts identified in the *cre1* and *xyr1* groups. The sequences related to enzymes and coexpressed with *cre1* (**a**) and *xyr1* (**b**) were annotated according to the main KO functions for Th3844, Th0179 and Ta0020. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844; KO: Kyoto Encyclopedia of Genes and Genomes Orthology.

In relation to the pathway classes triggered along with *cre1*, carbohydrate metabolism and amino acid metabolism were highly activated for Th3844 and Th0179, respectively, whereas for Ta0020, xenobiotic biodegradation and metabolism, nucleotide metabolism and carbohydrate metabolism were highly activated (**Fig. 3a**). The pathway classes enriched in the *xyr1* groups were identified as nucleotide metabolism and metabolism of cofactors and vitamins for Th3844, nucleotide metabolism for Th0179 and xenobiotic biodegradation and metabolism, lipid metabolism and amino acid metabolism for Ta0020 (**Fig. 3b**).

The number of metabolic pathways for the *cre1* and *xyr1* groups shared among all strains was larger than that observed exclusively for each strain or between only two strains/species (Additional file 7: Fig. S1). In addition, compared to the other two strains, Ta0020 presented the highest number of exclusive pathway classes related to both TFs (Additional file 7: Fig. S1).

Differentially Expressed Transcripts from *xyr1* and *cre1* Groups

We determined which of the transcripts present in the *cre1* and *xyr1* groups were differentially expressed considering the statistical analyses performed by [55]. We found different quantities of differentially expressed transcripts for Ta0020 (78) and Th0179 (6) (Additional file 9: Table S9). For Ta0020, most of the upregulated transcripts under cellulose growth (70) were associated with the *cre1* group, including *cre1* (*p-value* = 0), whereas *xyr1* showed a potentially significant expression level under glucose growth (*p-value* = 0.05) (Additional file 7: Table S10). However, for Th3844, *cre1* showed a significant modulation in expression (*p-value* = 0.03) between the two carbon sources, with a high expression level under glucose growth (Additional file 7: Table S10). Even with the reduced differentiation shown (-1.13), this phenomenon has already been considered by other studies [76-78]. For Th0179, *xyr1* (*p-value* = 0.08) and *cre1* (*p-value* = 0.27) showed no significant modulation in transcript expression between growth under cellulose or glucose, but *xyr1* had a higher transcript expression level under glucose, whereas *cre1* had a higher transcript expression level under cellulose growth conditions (Additional file 7: Table S10). For Th3844, the same was observed for *xyr1* (*p-value* = 0.29) but with a higher transcript expression level under glucose (Additional file 7: Table S10).

For Ta0020, the differentially expressed transcripts identified under cellulose growth in the *cre1* group included transcripts encoding CAZymes, sugar transporters, uncharacterized proteins and proteins related to secondary metabolism. Under glucose growth conditions, few differentially expressed transcripts were found. The differentially expressed transcripts identified in the *xyr1* group for Th0179 were not described yet as related to cellulose degradation (Additional file 9: Table S9).

CAZymes, TFs and Transporters Distribution Among *cre1* and *xyr1* Groups

Due to the importance of CAZymes, TFs and transporters for plant biomass degradation, we sought transcripts encoding these types of proteins within the *cre1* and

xyr1 groups (**Fig. 4**). Within the *cre1* group, Ta0020 presented the highest number of transcripts encoding these proteins (**Fig. 4a**). Between the *T. harzianum* strains, only Th0179 showed transcripts encoding TFs, transporters and CAZymes, whereas Th3844 presented transcripts encoding only CAZymes (**Fig. 4a**).

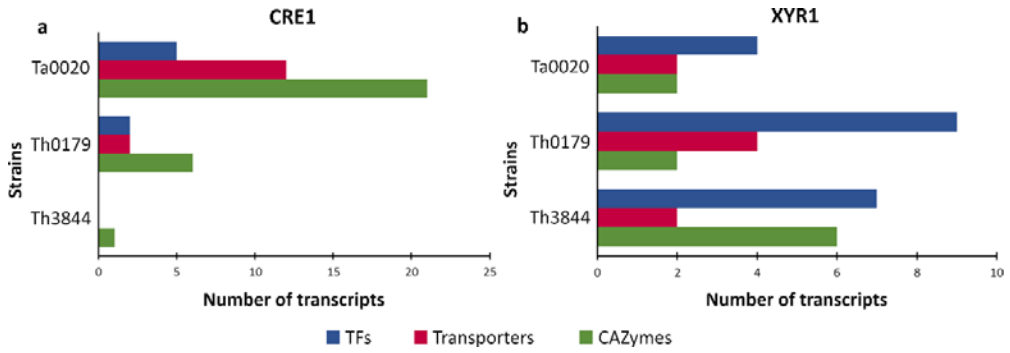

Fig. 4. TFs, transporters and CAZymes comparison among strains and groups. Number of transcripts encoding TFs, transporters and CAZymes for Ta0020, Th0179 and Th3844 distributed among *cre1* (**a**) and *xyr1* (**b**) groups. TFs: Transcription factors; CAZymes: Carbohydrate-active enzymes. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844.

The number of transcripts encoding TFs, transporters and CAZymes associated with *xyr1* were more distributed among all strains than those observed in the *cre1* group (**Fig. 4b**). However, compared with the other two strains, Th0179 showed the highest number of TFs and transporters. Th3844 presented the highest number of coexpressed CAZymes (**Fig. 4b**).

We also investigated transcripts encoding TFs, transporters and CAZymes within the *cre1* and *xyr1* subgroups for all analyzed strains. Within the *cre1* subgroup, Th3844 did not present transcripts encoding these types of proteins, whereas Th0179 presented transcripts encoding CAZymes, TFs and transporters, and Ta0020 showed transcripts encoding TFs and transporters. This profile was not observed in the *xyr1* subgroup, where Th3844 presented transcripts encoding CAZymes, TFs and transporters; Th0179 presented transporters and TFs; and Ta0020 presented a TF and a CAZyme.

CAZymes

Considering that hydrolytic enzymes play an important role in plant biomass degradation, a deep investigation of the CAZymes present in the *xyr1* and *cre1* groups was performed considering all strains. We identified the main CAZyme classes, such as carbohydrate esterases (CEs), glycoside hydrolases (GHs), glycosyltransferases (GTs) and polysaccharide lyases (PLs), in these groups, and their profiles were compared to detect similarities and differences between the strains (**Fig. 5**).

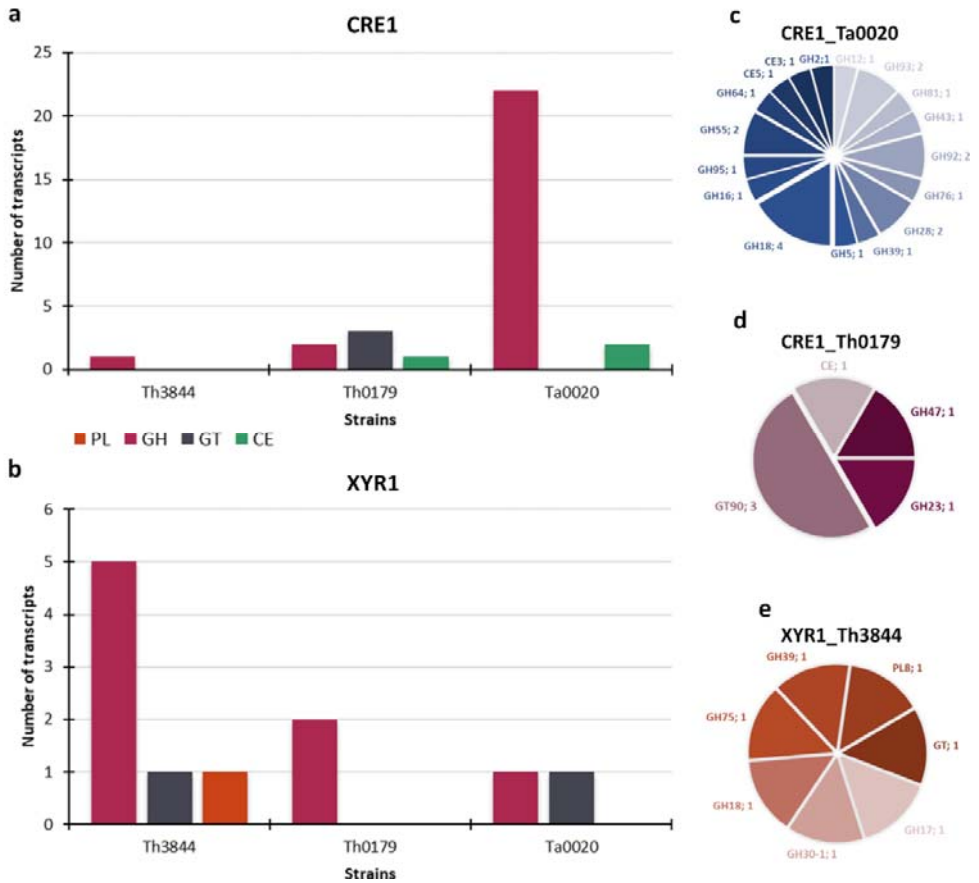

Fig. 5. Distribution of CAZyme families within *cre1* and *xyr1* groups for *Trichoderma* spp. Classification of CAZyme families coexpressed with *cre1* (**a**) and *xyr1* (**b**) for Th3844, Th0179 and Ta0020 and quantification of each CAZyme family in the *cre1* group for Ta0020 (**c**) and Th0179 (**d**) and in the *xyr1* group for Th3844 (**e**). PL: Polysaccharide lyases; GH: Glycoside hydrolases; GT: Glycosyltransferases; CE: Carbohydrate esterases; Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844.

For the *xyr1* and *cre1* groups, the GH family was identified in all evaluated strains, with different numbers (**Fig. 5a, b**). These enzymes constitute the main catalytic

machinery for the breakdown of glycosidic bonds between carbohydrate molecules [79]. Several GHs were also identified during the mycoparasitism process, whose activities can be related to fungal cell wall degradation in mycoparasite species of *Trichoderma* [80]. Most GHs related to mycoparasitic activity were associated with *cre1* in Ta0020 (**Fig. 5c**), which presented the highest number of classified transcripts from the GH family for the *cre1* group (**Fig. 5a**). Whereas Th0179 showed transcripts encoding GHs (**Fig. 5d**), Th3844 presented only one encoding GH28 coexpressed with *cre1*. For the *xyr1* group, this last strain presented a higher number of transcripts from the GH family (**Fig. 5e**) than did Th0179 (GH16 and GH76) and Ta0020 (GH75).

In addition to GHs, the CE family is also important for the efficient enzymatic hydrolysis of xylan [81] and was associated with the *cre1* transcript in Ta0020 (**Fig. 5c**) and Th0179 (**Fig. 5d**). The PL family, which cleaves certain activated glycosidic linkages in acidic polysaccharides [82], was coexpressed with the *xyr1* transcript in only Th3844 (**Fig. 5e**). In the *cre1* group, the GT family, which includes enzymes that act in the synthesis of the cell wall of microorganisms by the transference of chemical groups [83], was present in only Th0179 (**Fig. 5a**), while in the *xyr1* group, it was present in Th3844 (**Fig. 5e**) and Ta0020 (GT22).

To obtain more information about the hydrolytic enzymes associated with CRE1 and XYR1, we looked out for transcripts encoding CAZymes in the *cre1* and *xyr1* subgroups. Within the *cre1* subgroup, Ta0020 and Th3844 did not show transcripts encoding CAZymes, but we identified GT90 (THAR02_09987), GH23 (THAR02_00630) and CE1 (THAR02_01920) transcripts in Th0179. In the *xyr1* subgroup, Ta0020 presented GT22 (TRIATDRAFT_129991), and Th3844 presented GH18 (THAR02_04791), GH75 (THAR02_05160), PL8 (THAR02_00385) and GT (THAR02_10636) transcripts. Th0179 did not present transcripts encoding CAZymes in this subgroup.

Transcription Factors

Considering that the induction or repression of the genes encoding CAZymes, transporters and catabolic pathway enzymes requires the release of monosaccharides and the action of many TFs [40, 47], we sought transcripts encoding these regulators coexpressed with *cre1* and *xyr1*. Zn_2Cys_6 was the most abundant coexpressed for the strains that showed transcripts encoding TFs in the *cre1* and *xyr1* groups. In the *cre1* group for Ta0020, the number of transcripts encoding Zn_2Cys_6 was higher than that

observed for Th0179 (**Fig. 6a**). For the *xyr1* group, both Th3844 and Th0179 produced many more transcripts encoding Zn_2Cys_6 than Ta0020 (**Fig. 6b**). Most Zn_2Cys_6 proteins also contain a fungal-specific TF domain, known as the middle homology region, that has a proposed role in regulation [84]. We found transcripts encoding this fungal domain for Ta0020 (**Fig. 6a**) and Th3844 (**Fig. 6b**) in the *cre1* and *xyr1* groups, respectively.

Transcripts encoding C_2H_2, a TF that was already described to be involved in plant cell wall depolymerization for filamentous fungi [46], were identified in the *cre1* group for Ta0020 (**Fig. 6a**) and in the *xyr1* group for *T. harzianum* strains (**Fig. 6b**). Furthermore, we found a transcript encoding the SteA regulator, which is present in nearly all filamentous ascomycetes associated with the development and/or morphology of fungi [85], only in the *xyr1* group for Th0179 (**Fig. 6b**).

Considering that the gene regulation involved in plant biomass degradation is crucial, we also investigated transcripts encoding TFs in the *cre1* and *xyr1* subgroups. In the *cre1* subgroup, Ta0020 showed transcripts encoding Zn_2Cys_6 (TRIATDRAFT_182375), C_2H_2 (TRIATDRAFT_36276), and a fungal-specific TF domain (TRIATDRAFT_258080). For Th0179, we only found a transcript encoding Zn_2Cys_6 (THAR02_10668). Within the *xyr1* subgroup, transcripts encoding Zn_2Cys_6 were found for Th0179 (THAR02_01165 and THAR02_07523) and for Th3844 (THAR02_02597), the latter of which also presented a C_2H_2 transcript (THAR02_02590).

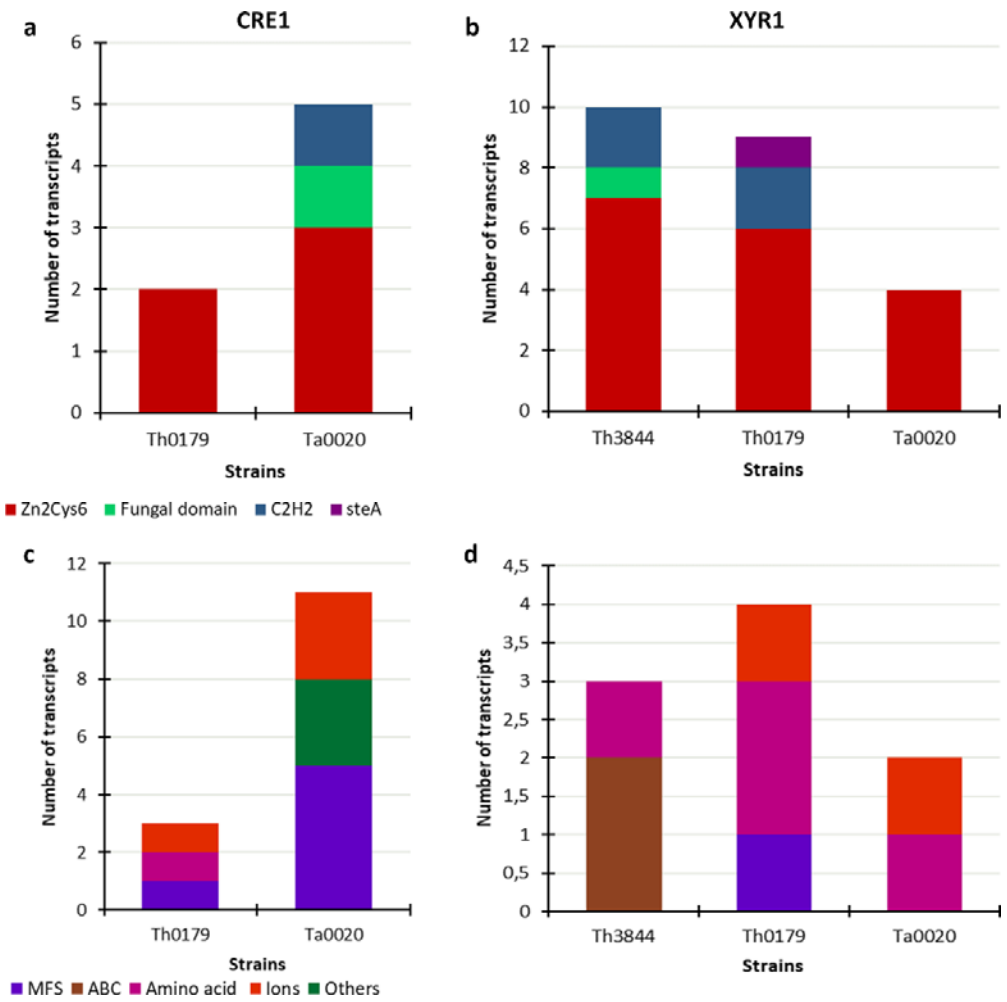

Fig. 6. Distribution of TFs and transporters within *cre1* and *xyr1* groups for *Trichoderma* spp. Classification of TFs coexpressed with *cre1* (**a**) and *xyr1* (**b**) and of transporters coexpressed with *cre1* (**c**) and *xyr1* (**d**) for Th3844, Th0179 and Ta0020. Th3844: *T. harzianum* IOC-3844; Th0179: *T. harzianum* CBMAI-0179; Ta0020: *T. atroviride* CBMAI-0020.

Transporters

Transporters are also important players in the degradation of plant biomass. It is well known that approximately 5% of the *T. reesei* genome comprises genes that encode proteins involved in transport [66]. Thus, we sought transcripts encoding transport proteins in the *cre1* and *xyr1* groups. For the *cre1* group, Th3844 did not coexpress transcripts encoding transporters (**Fig. 6c**). On the other hand, this strain showed transcripts with a similar expression pattern as *xyr1* (**Fig. 6d**). These transcripts encode ATP-binding cassette (ABC) transporters, which act as ATP-driven transporters for

several substrates, including lipids, sugars, drugs, heavy metals, and auxin [49], and amino acid transporters, which probably constitute a major nutritional source for all fungi [86]. In addition, all strains presented transcripts encoding amino acid transporters coexpressed with *xyr1* in different amounts (**Fig. 6d**). For the *cre1* group, only Th0179 showed transcripts encoding this type of transport protein.

Another important class of protein transport involved in plant biomass degradation is major facilitator superfamily (MFS) permeases. In filamentous fungi, these proteins transport a wide range of substrates, including small molecule inducers from the extracellular environment into the fungus, which could influence the expression of CAZyme-encoding genes [78]. In the *cre1* group, MFS were the most abundant proteins among the transporters for Ta0020 (**Fig. 6c**). For the *xyr1* group, only Th0179 showed transcripts encoding this type of protein transport (**Fig. 6d**).

Transcripts encoding ion transporters were present in the *cre1* group (**Fig. 6c**), including transcripts for zinc-regulated transporter 1 (TRIATDRAFT_147496) and a magnesium ion transporter (TRIATDRAFT_315729) for Ta0020 and for potassium transporter 8 (THAR02_02027) for Th0179. In the same way, transcripts encoding ion transporters were also present in the *xyr1* group (**Fig. 6d**), such as transcripts for a $Cd^{2+}Zn^{2+}$ transporter (TRIATDRAFT_320571) for Ta0020 and for a Ca^{2+} transporter (THAR02_01350) for Th0179.

In relation to the *cre1* subgroup, for Ta0020, we found transcripts encoding an MFS permease transporter (TRIATDRAFT_89305 and TRIATDRAFT_317997), whereas for Th0179, a transcript encoding a potassium transporter (THAR02_02027) was found. In the *xyr1* subgroup, transcripts encoded a $Cd^{2+}Zn^{2+}$ transporter (TRIATDRAFT_320571) for Ta0020, an amino acid transporter (THAR02_07655) and a Ca^{2+} transporter (THAR02_01350) for Th0179 and ABC transporters (THAR02_04735 and THAR02_04149) for Th3844.

Discussion

Due to the ability to mycoparasitize other fungi, *T. atroviride* and *T. harzianum* are widely used in agriculture as biocontrol agents against phytopathogens [87-89]. *T. harzianum* has also demonstrated considerable enzymatic potential for application in lignocellulosic biomass hydrolysis [27, 90, 91]. However, the transcriptional regulation involved in the expression of CAZyme-encoding genes in this species is still poorly explored when compared to that of the major producer of cellulases, *T. reesei*. In this

study, we chose to investigate and compare the genetic mechanisms involved in XYR1- and CRE1-mediated gene regulation in *Trichoderma* spp. For this, we used transcriptome data [55] for Th3844 and Th0179, both mycoparasite strains with hydrolytic potential, and Ta0020, a common mycoparasite ancestor, to model the gene coexpression networks for each. We used this RNA-Seq dataset since a difference in gene expression among these three strains was established, showing that it was possible to capture differences among them [55].

The diversification observed for CRE1 and XYR1 inside Ascomycota may indicate a high functional diversity for these proteins, which can reflect changes in their functions [92]. Furthermore, for both regulators, a differentiation process was observed for Ta0020 compared to Th3844 and Th0179 among *Trichoderma* spp. These phylogenetic relationships may be related to the different types of enzymatic performance previously reported, in which Ta0020 had a lower cellulase activity profile during growth on cellulose compared to *T. harzianum* strains [55].

Analyzing the gene coexpression data using a clustering method allowed us to identify correlations among genes with similar expression patterns, which often indicates that they are active in the same biological processes [51, 52, 93]. Our gene coexpression network analysis revealed that *T. harzianum* and *T. atroviride* have developed different strategies related to gene regulation involved in cellulose and hemicellulose degradation, which were already described for other fungi [41, 76, 94-96]. We suggest that since Th3844 and Th0179 had a larger number of transcripts associated with *xyr1* than with *cre1*, there were also more metabolic pathways related to XYR1, the main cellulase and hemicellulase activator. At a more specific level, the coexpression network topology of the subgroups revealed that the three strains presented different profiles of coexpressed transcripts with *xyr1* and *cre1*, which were more accentuated in *T. atroviride*.

Assuming that hubs are frequently more relevant to the functionality of networks than other nodes are [47], they were sought out within the *cre1* and *xyr1* groups, and new targets that were coexpressed with the main regulators involved in cellulose degradation were discovered. Although transcripts encoding uncharacterized proteins were identified as potential hub nodes in the *xyr1* groups, indicating their importance in the degradation process, characterized proteins were found in the *cre1* groups. Transcripts encoding proteins related to the proteasome, which is a highly sophisticated protease complex responsible for the degradation of the majority of intracellular

proteins [97]; oxidative phosphorylation, which is responsible for mitochondrial ATP production related to the energy required for cell growth, differentiation, and development [98]; and nucleoside import into fungal cells, which can be associated with nucleotide synthesis via salvage pathways in cells lacking the capacity for *de novo* synthesis [99], were found to be hubs in Th3844, Th0179 and Ta0020, respectively.

For *Trichoderma* spp., GO enrichment in the *cre1* group showed that terms related to lignocellulose breakdown and to regulation of gene expression were the main enriched terms. For the *xyr1* group, all these terms or correlated terms were also enriched, except for hydrolase activity, which was present in only Th3844. In a recent study, it was reported that several terms related to lignocellulose breakdown were enriched in the *xyr1* group for *T. reesei* RUT-C30 grown on bagasse or fructose as a carbon source [53]. Furthermore, XYR1 and CRE1 are zinc finger proteins, which encompass a wide variety of compact protein domains stabilized by a structural zinc ion [100]. Several classes of zinc finger motifs are present in TFs and act as parts of DNA-binding and protein-protein interaction domains [101]. In the *cre1* and *xyr1* groups, we found transcripts enriched with metal ion binding and zinc ion binding, respectively. In this way, we can suppose that transcripts encoding zinc finger proteins were coexpressed with them, as we found.

To obtain more information about the metabolic pathways triggered by CRE1 and XYR1 in *Trichoderma* spp., we investigated the enzymatic activity in the *cre1* and *xyr1* groups. Our data showed high similarity between metabolic pathways triggered by both regulators during cellulose degradation. However, comparing Ta0020 and Th0179, they had distinct metabolic processes occurring within the *cre1* and *xyr1* groups, which suggests that these two TFs have developed different functional regulation in each strain. In addition, the exclusive pathway classes for Ta0020 were not related to the biomass degradation process. For instance, the metabolism of terpenoids and polyketides, two secondary metabolite compounds involved in the antibiosis process during mycoparasitic activity [102, 103], was identified.

We also investigated the expression profile of the transcripts within the *cre1* and *xyr1* groups, including *xyr1* and *cre1*, for all evaluated strains under cellulose and glucose growth. In 2016, Castro et al. (2016) observed that in the presence of glucose, the expression level of the transcript *xyr1* for *T. reesei* was minimal [76]. Our findings did not show this for Ta0020. In this case, the *xyr1* expression level was higher under glucose growth than under cellulose growth. In addition, *cre1* was upregulated in the

presence of cellulose, which was not expected due to the role of CRE1 in the repression of cellulolytic and hemicellulolytic enzymes when an easily metabolizable sugar, i.e., glucose, is available in the environment [46]. Conversely, the *cre1* expression level was higher in the presence of glucose than in the presence of cellulose for Th3844. Upregulated transcripts within the *cre1* group for Ta0020 were also not expected. Here, it is important to mention that the gene coexpression networks were modeled based on the transcript expression level under two sets of conditions. Considering that the expression of the *cre1* transcript for Ta0020 was upregulated under cellulose growth, the expression of other transcripts in the *cre1* group was also expected to be upregulated under these conditions.

We also investigated the distribution of transcripts encoding CAZymes, TFs, and transporters within the *cre1* and *xyr1* groups among *Trichoderma* spp. Initially, it was believed that CRE1 directly repressed the transcription of cellulase and hemicellulase genes when a readily metabolizable carbon source was present in the medium [43]. While this remains correct, studies in *T. reesei* have demonstrated that CRE1 may also control genes related to several other functions, and only a small number of CAZyme genes are direct targets of this TF during CCR [77, 104, 105]. In this way, we did not expect to find transcripts encoding CAZymes in *cre1* groups. Herein, we had only a few transcripts encoding CAZymes in *cre1* groups for Th3844 and Th0179, whereas others encode proteins with diverse functions. In relation to TFs and transporters, Antoniêto et al. (2014) observed that transcripts encoding these proteins were strongly repressed by CRE1 [77] in *T. reesei*. Thus, we expected to find a large number of transcripts encoding TFs and transporters coexpressed with the *cre1* transcript for all strains, but it was observed only for Ta0020, especially for the transcripts encoding proteins related to transport.

Castro et al. (2016) demonstrated for *T. reesei* that the expression of CAZymes, TFs, and transporters was affected by the TF XYR1 in the context of plant biomass degradation [76]. Furthermore, Reithner et al. (2014) suggested that XYR1 was also involved in the mediation of the plant-fungus recognition processes in *T. atroviride* [45]. It was also reported that the role of XYR1 differs strongly among ascomycete fungi [44]. We found transcripts encoding CAZymes, TFs and transporters coexpressed with *xyr1* for the three analyzed strains. Although these transcripts were present in Ta0020, they may induce different mechanisms there than in Th3844 and Th0179.

Our results showed that *T. atroviride* has an increased number of GHs coexpressed with *cre1* that are related to mycoparasitism and not to cellulose and hemicellulose degradation. For instance, GH18, whose members have chitinase activity, is the largest family of CAZymes in mycoparasitic *Trichoderma* species [80], whereas GH28, which is the main family involved in the degradation of pectin [106], is also involved in mycoparasitism. Moreover, β-glucanases have been described for mycoparasitic species within *Trichoderma* spp. in GH families 55, 64 and 81 [58]. Among the CAZymes responsible for cellulose degradation, we found transcripts encoding only GH5 and GH12, and for hemicellulose degradation, we found transcripts for only CE5, GH95 and GH43 [31].

Almeida et al. (2020) analyzed the main CAZyme classes detected in the transcriptome of Th3844, Th0179 and Ta0020 under cellulose or glucose growth conditions. They found the main CAZyme families responsible for cellulose and hemicellulose degradation for all analyzed strains [55]. Herein, the transcripts encoding cellulases and hemicellulases were not found in the *cre1* and *xyr1* groups for *T. harzianum*. It is possible that the transcripts encoding these hydrolytic enzymes were grouped in other formed groups within the coexpression networks, mainly because their expression patterns were different from those observed for the transcripts encoding XYR1 and CRE1.

The production of cellulases and hemicellulases in *T. reesei* occurs by activation of gene expression via specific inducers in balance with the repression of gene expression via CCR [47]. The sensing of inducers starts a signaling pathway resulting in the activation of TFs, which is followed by the production of plant biomass-degrading enzymes and the activation of metabolic pathways to use the accumulating sugar monomers [46, 107]. Transporters also have an important role in the utilization of lignocellulose by fungi, promoting the transport of molecules between the intracellular and extracellular spaces and, consequently, controlling CAZyme-encoding gene expression [108]. For *T. reesei*, MFS, amino acids and ABC transporters were modulated by XYR1 and CRE1 [76, 77, 104]. Our results suggest that the expression of transporters and TFs during cellulose degradation was affected by XYR1 in *T. harzianum*, as was demonstrated for *T. reesei* [76]. In addition, *T. harzianum* and *T. atroviride* developed different strategies related to the import of molecules from the environment that can influence the production of plant cell wall-degrading enzymes, which may result in contrasting enzymatic performances during cellulose degradation.

Conclusions

Our results propose that the coexpressed transcripts with *xyr1* and *cre1* are different for each strain and suggest that different strategies related to XYR1 and CRE1 are used by *Trichoderma* spp. during cellulose degradation and glucose metabolization. It was observed that the transcripts encoding CAZymes related to cellulose and hemicellulose degradation in *T. harzianum* are not necessarily coexpressed with the TFs studied herein. Taken together, our findings provide new insights into transcripts coexpressed with XYR1 and CRE1 in *T. harzianum* strains, which could be the target of new studies to evaluate their role in lignocellulose degradation. Finally, our results can contribute to a better understanding of the genetic mechanisms involved in hydrolytic enzyme regulation for mycoparasite *Trichoderma* species. This will be important to expand the potential of *T. harzianum* not only in the development of better enzyme cocktails but also in many other industrial applications.

Additional files

Additional file 1: Table S1. Description of the species used for the phylogenetic analysis of the transcription factors CRE1 and XYR1 in Ascomycota.

Additional file 2: Table S2. Filtered transcripts used to model the gene coexpression networks among *Trichoderma* spp.

Additional file 3: Table S3. Transcripts identified in the cluster analysis of gene coexpression networks of *Trichoderma* spp.

Additional file 4: Table S4. Transcripts identified as coexpressed with *cre1* and *xyr1* in gene coexpression networks of *Trichoderma* spp.

Additional file 5: Table S5. Subgroups identified among *cre1* and *xyr1* groups using the MCODE methodology for all evaluated strains.

Additional file 6: Table S6. Subgroups identified among the *cre1* and *xyr1* groups using the Louvain algorithm for all evaluated strains.

Additional file 7: Table S7. Number of transcripts coexpressed in each subgroup formed from *cre1* and *xyr1* groups for *Trichoderma* spp. and its corresponding color. **Fig. S1.** Venn diagrams of pathway classes shared between *Trichoderma* spp. with expression levels associated with *cre1* (**a**) and *xyr1* (**b**) transcript levels. **Table 10.** Expression values in TPM of *cre1* and *xyr1* for *Trichoderma* spp. under cellulose and glucose growth conditions.

Additional file 8: Table S8. Hubs identified within the *cre1* and *xyr1* groups of *Trichoderma* spp.

Additional file 9: Table S9. Differentially expressed transcripts established in the *cre1* and *xyr1* groups for all evaluated strains.

List of Abbreviations

1G ethanol: First-Generation Ethanol; **2G ethanol:** Second-Generation Ethanol; **ABC:** ATP-binding cassette; **CAZymes:** Carbohydrate-Active Enzymes; **CBMAI:** Brazilian Collection of Environment and Industry Microorganisms; **CBMEG:** Center for Molecular Biology and Genetic Engineering; **CCR:** Carbon Catabolite Repression; **CEs:** Carbohydrate Esterases; **CNT:** Concentrative Nucleoside Transporter; **CPQBA:** Chemical, Biological and Agricultural Pluridisciplinary Research Center; **CRE1:** Carbon Catabolite Repressor 1; **GHs:** Glycoside Hydrolases; **GO:** Gene Ontology; **GTs:** Glycosyltransferases; **JTT:** Jones-Taylor-Thornton; **KEGG:** Kyoto Encyclopedia of Genes and Genomes; **KOs:** KEGG Orthologies; **LPMOs:** Lytic Polysaccharides Monooxygenases; **MCODE**: Molecular Complex Detection; **MEGA:** Molecular Evolutionary Genetics Analysis; **MFS:** Major Facilitator Superfamily; **MRP:** Mitochondrial Ribosomal Protein; **ML:** Maximum Likelihood; **NCBI:** National Center for Biotechnology Information; **PLs:** Polysaccharide Lyases; **RNA-Seq:** RNA Sequencing; **Ta0020:** *Trichoderma atroviride* CBMAI-0020; **TFs:** Transcription Factors; **Th0179:** *Trichoderma harzianum* CBMAI-0179; **Th3844:** *Trichoderma harzianum* IOC-3844; **TPM:** Transcripts Per Million; **UNICAMP:** University of Campinas; **WGCNA:** Weighted Gene Coexpression Network Approach; **XYR1:** Xylanase Regulator 1.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets generated and/or analyzed during this study are included in this published article as Additional files 1, 2, 3, 4, 5, 6, 7, 8 and 9.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2015/09202-0, 2018/19660-4), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Computational Biology Program and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). RRR received a partial MS fellowship from the CAPES Computational Biology Program (88887.176241/2018-00) and partial support from CAPES – PROEX (88882.329483/2019-01); AHA received a PhD from the FAPESP (2019/03232-6); DAA received a partial MS fellowship from the FAPESP (2017/17782-2) and partial support from the CAPES Computational Biology Program (88887.336686/2019-00); JAFF received a PhD fellowship from the CNPq (170565/2017-3) and a PhD fellowship from the CAPES (CBP – 88887.334235/2019-00); MACH received a PD fellowship from FAPESP (2018/18856-1); and APS is the recipient of a research fellowship from the CNPq.

Author contributions

RRR and APS conceived and designed the study. JAFF contributed to the study performance. RRR, AHA and DAA performed the data analysis. MACH designed and achieved the fermentation experiments. RRR and AHA drafted the manuscript, which was critically revised by DAA, JAFF and APS. All authors read and approved the final manuscript.

Acknowledgments

We would like to acknowledge funding from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2015/09202-0 and 2018/19660-4), the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Computational Biology Program) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We thank 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 the center and laboratory space.

References

1. Gupta A, Verma JP: **Sustainable bio-ethanol production from agro-residues: A review.** *Renewable and Sustainable Energy Reviews* 2015, **41:**550-567.
2. Kucharska K, Słupek E, Cieśliński H, Kamiński M: **Advantageous conditions of saccharification of lignocellulosic biomass for biofuels generation via fermentation processes.** *Chemical Papers* 2019.
3. Sharma HK, Xu C, Qin W: **Biological Pretreatment of Lignocellulosic Biomass for Biofuels and Bioproducts: An Overview.** *Waste and Biomass Valorization* 2019, **10:**235-251.
4. Association RF: **Renewable Fuels Association.** 2019.
5. Alonso Pippo W, Luengo CA, Alonsoamador Morales Alberteris L, Garzone P, Cornacchia G: **Energy Recovery from Sugarcane-Trash in the Light of 2nd Generation Biofuels. Part 1: Current Situation and Environmental Aspects.** *Waste and Biomass Valorization* 2011, **2:**1-16.
6. Sun JX, Sun XF, Sun RC, Su YQ: **Fractional extraction and structural characterization of sugarcane bagasse hemicelluloses.** *Carbohydrate Polymers* 2004, **56:**195-204.
7. Méndez J, Passos DdF, Wischral D, Modesto LF, Pereira N: **Second-generation ethanol production by separate hydrolysis and fermentation from sugarcane bagasse with cellulose hydrolysis using a customized enzyme cocktail.** *Biofuels* 2019**:**1-7.
8. Robak K, Balcerek M: **Review of Second Generation Bioethanol Production from Residual Biomass.** *Food technology and biotechnology* 2018, **56:**174-187.
9. Ruchala J, Kurylenko OO, Dmytruk KV, Sibirny AA: **Construction of advanced producers of first- and second-generation ethanol in Saccharomyces cerevisiae and selected species of non-conventional yeasts**

(Scheffersomyces stipitis, Ogataea polymorpha). *Journal of Industrial Microbiology & Biotechnology* 2019.

10. Ferreira JA, Brancoli P, Agnihotri S, Bolton K, Taherzadeh MJ: **A review of integration strategies of lignocelluloses and other wastes in 1st generation bioethanol processes.** *Process Biochemistry* 2018, **75:**173-186.
11. Alvira P, Negro MJ, Ballesteros M: **Effect of endoxylanase and α-l-arabinofuranosidase supplementation on the enzymatic hydrolysis of steam exploded wheat straw.** *Bioresource Technology* 2011, **102:**4552-4558.
12. Le Costaouëc T, Pakarinen A, Várnai A, Puranen T, Viikari L: **The role of carbohydrate binding module (CBM) at high substrate consistency: Comparison of Trichoderma reesei and Thermoascus aurantiacus Cel7A (CBHI) and Cel5A (EGII).** *Bioresource Technology* 2013, **143:**196-203.
13. Bomble YJ, Lin C-Y, Amore A, Wei H, Holwerda EK, Ciesielski PN, Donohoe BS, Decker SR, Lynd LR, Himmel ME: **Lignocellulose deconstruction in the biosphere.** *Current Opinion in Chemical Biology* 2017, **41:**61-70.
14. Sartori T, Tibolla H, Prigo E, Colla LM, Costa JAV, Bertolin TE: **Enzymatic saccharification of lignocellulosic residues by cellulases obtained from solid state fermentation using Trichoderma viride.** *BioMed Research International* 2015.
15. Peciulyte A, Pisano M, de Vries RP, Olsson L: **Hydrolytic potential of five fungal supernatants to enhance a commercial enzyme cocktail.** *Biotechnology letters* 2017, **39:**1403-1411.
16. Zafra G, Cortés-Espinosa DV: **Biodegradation of polycyclic aromatic hydrocarbons by Trichoderma species: a mini review.** *Environmental Science and Pollution Research* 2015, **22:**19426-19433.
17. Bischof RH, Ramoni J, Seiboth B: **Cellulases and beyond: the first 70 years of the enzyme producer Trichoderma reesei.** *Microbial Cell Factories* 2016, **15:**106.
18. Sridevi A, Ramanjaneyulu G, Suvarnalatha Devi P: **Biobleaching of paper pulp with xylanase produced by Trichoderma asperellum.** *3 Biotech* 2017, **7:**266-266.
19. Kumar M, Ashraf S: **Role of Trichoderma spp. as a Biocontrol Agent of Fungal Plant Pathogens.** In *Probiotics and Plant Health.* Edited by Kumar V, Kumar M, Sharma S, Prasad R. Singapore: Springer Singapore; 2017: 497-506

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

21. Mukherjee AK, Sampath Kumar A, Kranthi S, Mukherjee PK: **Biocontrol potential of three novel Trichoderma strains: isolation, evaluation and formulation.** *3 Biotech* 2014, **4:**275-281.

22. Atanasova L, Crom SL, Gruber S, Coulpier F, Seidl-Seiboth V, Kubicek CP, Druzhinina IS: **Comparative transcriptomics reveals different strategies of Trichodermamycoparasitism.** *BMC Genomics* 2013, **14:**121.

23. Druzhinina IS, Kopchinskiy AG, Kubicek CP: **The first 100 Trichoderma species characterized by molecular data.** *Mycoscience* 2006, **47:**55.

24. Chaverri P, Samuels GJ, Stewart EL: **Hypocrea virens sp. nov., the Teleomorph of Trichoderma virens.** *Mycologia* 2001, **93:**1113-1124.

25. Dodd SL, Lieckfeldt E, Samuels GJ: **Hypocrea atroviridis sp. nov., the teleomorph of Trichoderma atroviride.** *Mycologia* 2003, **95:**27-40.

26. Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J: **The Trichoderma harzianum demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages.** *BMC Evolutionary Biology* 2010, **10:**94.

27. Horta MAC, Vicentini R, Delabona PdS, Laborda P, Crucello A, Freitas S, Kuroshu RM, Polikarpov I, Pradella JGdC, Souza AP: **Transcriptome profile of Trichoderma harzianum IOC-3844 induced by sugarcane bagasse.** *PloS one* 2014, **9:**e88689-e88689.

28. Li J-X, Zhang F, Li J, Zhang Z, Bai F-W, Chen J, Zhao X-Q: **Rapid production of lignocellulolytic enzymes by Trichoderma harzianum LZ117 isolated from Tibet for biomass degradation.** *Bioresource technology* 2019, **292:**122063.

29. Santos CA, Ferreira-Filho JA, O'Donovan A, Gupta VK, Tuohy MG, Souza AP: **Production of a recombinant swollenin from Trichoderma harzianum in Escherichia coli and its potential synergistic role in biomass degradation.** *Microbial Cell Factories* 2017, **16:**83.

30. Santos CA, Morais MAB, Terrett OM, Lyczakowski JJ, Zanphorlin LM, Ferreira-Filho JA, Tonoli CCC, Murakami MT, Dupree P, Souza AP: **An engineered GH1 β-glucosidase displays enhanced glucose tolerance and**

increased sugar release from lignocellulosic materials. *Scientific Reports* 2019, **9:**4903.

31. Ferreira Filho JA, 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.
32. Crucello A, Sforça DA, Horta MAC, dos Santos CA, Viana AJC, Beloti LL, de Toledo MAS, Vincentz M, Kuroshu RM, de Souza AP: **Analysis of Genomic Regions of Trichoderma harzianum IOC-3844 Related to Biomass Degradation.** *PLOS ONE* 2015, **10:**e0122122.
33. Binod P, Gnansounou E, Sindhu R, Pandey A: **Enzymes for second generation biofuels: Recent developments and future perspectives.** *Bioresource Technology Reports* 2019, **5:**317-325.
34. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen J-CN, Johansen KS, Krogh KBRM, Jørgensen CI, Tovborg M, Anthonsen A, et al: **Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components.** *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108:**15079-15084.
35. Saloheimo M, Paloheimo M, Hakola S, Pere J, Swanson B, Nyyssönen E, Bhatia A, Ward M, Penttilä M: **Swollenin, a Trichoderma reesei protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials.** *European Journal of Biochemistry* 2002, **269:**4202-4211.
36. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: **The carbohydrate-active enzymes database (CAZy) in 2013.** *Nucleic acids research* 2014, **42:**D490-D495.
37. Castro LdS, Antoniêto ACC, Pedersoli WR, Silva-Rocha R, Persinoti GF, Silva RN: **Expression pattern of cellulolytic and xylanolytic genes regulated by transcriptional factors XYR1 and CRE1 are affected by carbon source in Trichoderma reesei.** *Gene Expression Patterns* 2014, **14:**88-95.
38. Amore A, Giacobbe S, Faraco V: **Regulation of cellulase and hemicellulase gene expression in fungi.** *Current genomics* 2013, **14:**230-249.
39. Dashtban M, Buchkowski R, Qin W: **Effect of different carbon sources on cellulase production by Hypocrea jecorina (Trichoderma reesei) strains.** *International journal of biochemistry and molecular biology* 2011, **2:**274-286.

40. Portnoy T, Margeot A, Seidl-Seiboth V, Le Crom S, Ben Chaabane F, Linke R, Seiboth B, Kubicek CP: **Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in Trichoderma reesei strains producing high and low levels of cellulase.** *Eukaryotic cell* 2011, **10:**262-271.

41. Stricker AR, Grosstessner-Hain K, Würleitner E, Mach RL: **Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in Hypocrea jecorina.** *Eukaryotic cell* 2006, **5:**2128-2137.

42. Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, Sándor E, Hartl L, Karaffa L, Druzhinina IS, Seiboth B, et al: **The CRE1 carbon catabolite repressor of the fungus Trichoderma reesei: a master regulator of carbon assimilation.** *BMC genomics* 2011, **12:**269-269.

43. Strauss J, Mach RL, Zeilinger S, Hartler G, Stöffler G, Wolschek M, Kubicek CP: **Crel, the carbon catabolite repressor protein from Trichoderma reesei.** *FEBS Letters* 1995, **376:**103-107.

44. Klaubauf S, Narang HM, Post H, Zhou M, Brunner K, Mach-Aigner AR, Mach RL, Heck AJR, Altelaar AFM, de Vries RP: **Similar is not the same: Differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi.** *Fungal Genetics and Biology* 2014, **72:**73-81.

45. Reithner B, Mach-Aigner AR, Herrera-Estrella A, Mach RL: **Trichoderma atroviride transcriptional regulator Xyr1 supports the induction of systemic resistance in plants.** *Applied and environmental microbiology* 2014, **80:**5274-5281.

46. Benocci T, Aguilar-Pontes MV, Zhou M, Seiboth B, de Vries RP: **Regulators of plant biomass degradation in ascomycetous fungi.** *Biotechnology for biofuels* 2017, **10:**152-152.

47. Alazi E, Ram AFJ: **Modulating Transcriptional Regulation of Plant Biomass Degrading Enzyme Networks for Rational Design of Industrial Fungal Strains.** *Frontiers in bioengineering and biotechnology* 2018, **6:**133-133.

48. Shida Y, Furukawa T, Ogasawara W: **Deciphering the molecular mechanisms behind cellulase production in Trichoderma reesei, the hyper-cellulolytic filamentous fungus.** *Bioscience, Biotechnology, and Biochemistry* 2016, **80:**1712-1729.

49. de Paula RG, Antoniêto ACC, Ribeiro LFC, Carraro CB, Nogueira KMV, Lopes DCB, Silva AC, Zerbini MT, Pedersoli WR, Costa MdN, Silva RN: **New Genomic Approaches to Enhance Biomass Degradation by the Industrial Fungus Trichoderma reesei.** *International Journal of Genomics* 2018, **2018:**17.

50. Zhang F, Bunterngsook B, Li J-X, Zhao X-Q, Champreda V, Liu C-G, Bai F-W: **Chapter Three - Regulation and production of lignocellulolytic enzymes from Trichoderma reesei for biofuels production.** In *Advances in Bioenergy. Volume* 4. Edited by Li Y, Ge X: Elsevier; 2019: 79-119

51. van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP: **Gene co-expression analysis for functional classification and gene-disease predictions.** *Briefings in bioinformatics* 2018, **19:**575-592.

52. Pirim H: **Construction of Gene Networks Using Expression Profiles.** In *Soft Computing for Biological Systems*. Edited by Purohit HJ, Kalia VC, More RP. Singapore: Springer Singapore; 2018: 67-89

53. Borin GP, Carazzolle MF, dos Santos RAC, Riaño-Pachón DM, Oliveira JVdC: **Gene Co-expression Network Reveals Potential New Genes Related to Sugarcane Bagasse Degradation in Trichoderma reesei RUT-30.** *Frontiers in Bioengineering and Biotechnology* 2018, **6**.

54. Horta MAC, Filho JAF, Murad NF, de Oliveira Santos E, dos Santos CA, Mendes JS, Brandão MM, Azzoni SF, de Souza AP: **Network of proteins, enzymes and genes linked to biomass degradation shared by Trichoderma species.** *Scientific Reports* 2018, **8:**1341.

55. Almeida DA, Horta MAC, Ferreira Filho JA, Murad NF, de Souza AP: **The synergistic actions of hydrolytic genes in coexpression networks reveal the potential of Trichoderma harzianum for cellulose degradation.** *bioRxiv* 2020**:**2020.2001.2014.906529.

56. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30:**2114-2120.

57. Baroncelli R, Piaggeschi G, Fiorini L, Bertolini E, Zapparata A, Pè ME, Sarrocco S, Vannacci G: **Draft Whole-Genome Sequence of the Biocontrol Agent Trichoderma harzianum T6776.** *Genome announcements* 2015, **3:**e00647-00615.

58. Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, et al: **Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of Trichoderma.** *Genome Biology* 2011, **12:**R40.

59. CLC: **Genomics Workbench (QIAGEN).** 6.5.2 edition. Aarhus, Denmark; 2016.

60. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic acids research* 1994, **22:**4673-4680.

61. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.** *Molecular Biology and Evolution* 2016, **33:**1870-1874.

62. Jones DT, Taylor WR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Bioinformatics* 1992, **8:**275-282.

63. Felsenstein J: **Confidence Limits on Phylogenies: An Approach Using the Bootstrap.** *Evolution* 1985, **39:**783-791.

64. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9:**559.

65. Team RC: **R: A language and environment for statistical computing.** In *R Foundation for Statistical Computing*. Vienna, Austria; 2018.

66. Langfelder P, Zhang B, Horvath S: **Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R.** *Bioinformatics* 2007, **24:**719-720.

67. Sanner MF: **Python: a programming language for software integration and development.** *Journal of molecular graphics & modelling* 1999, **17:**57-61.

68. Hagberg A, Swart P, S Chult D: **Exploring network structure, dynamics, and function using networkx.** In; *2008-01-01; United States*. 2008

69. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E: **Fast unfolding of communities in large networks.** *Journal of Statistical Mechanics: Theory and Experiment* 2008, **2008:**P10008.

70. Bader GD, Hogue CWV: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC Bioinformatics* 2003, **4:**2.

71. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene Ontology: tool for the unification of biology.** *Nature Genetics* 2000, **25:**25-29.

72. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic acids research* 2000, **28:**27-30.

73. Supek F, Bošnjak M, Škunca N, Šmuc T: **REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms.** *PLOS ONE* 2011, **6:**e21800.

74. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL: **Biopython: freely available Python tools for computational molecular biology and bioinformatics.** *Bioinformatics* 2009, **25:**1422-1423.

75. Mukaka MM: **Statistics corner: A guide to appropriate use of correlation coefficient in medical research.** *Malawi medical journal : the journal of Medical Association of Malawi* 2012, **24:**69-71.

76. dos Santos Castro L, de Paula RG, Antoniêto ACC, Persinoti GF, Silva-Rocha R, Silva RN: **Understanding the Role of the Master Regulator XYR1 in Trichoderma reesei by Global Transcriptional Analysis.** *Frontiers in Microbiology* 2016, **7**.

77. Antoniêto ACC, dos Santos Castro L, Silva-Rocha R, Persinoti GF, Silva RN: **Defining the genome-wide role of CRE1 during carbon catabolite repression in Trichoderma reesei using RNA-Seq analysis.** *Fungal Genetics and Biology* 2014, **73:**93-103.

78. Antoniêto ACC, de Paula RG, Castro LDS, Silva-Rocha R, Persinoti GF, Silva RN: **Trichoderma reesei CRE1-mediated Carbon Catabolite Repression in Re-sponse to Sophorose Through RNA Sequencing Analysis.** *Current genomics* 2016, **17:**119-131.

79. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.** *Nucleic acids research* 2009, **37:**D233-D238.

80. Gruber S, Seidl-Seiboth V: **Self versus non-self: fungal cell wall degradation in Trichoderma.** *Microbiology* 2012, **158:**26-34.

81. Malgas S, Mafa MS, Mkabayi L, Pletschke BI: **A mini review of xylanolytic enzymes with regards to their synergistic interactions during hetero-xylan degradation.** *World Journal of Microbiology and Biotechnology* 2019, **35:**187.

82. Linhardt RJ, Galliher PM, Cooney CL: **Polysaccharide lyases.** *Applied Biochemistry and Biotechnology* 1987, **12:**135-176.

83. Ardèvol A, Rovira C: **Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations.** *Journal of the American Chemical Society* 2015, **137:**7528-7547.

84. MacPherson S, Larochelle M, Turcotte B: **A fungal family of transcriptional regulators: the zinc cluster proteins.** *Microbiology and molecular biology reviews : MMBR* 2006, **70:**583-604.

85. Todd RB, Zhou M, Ohm RA, Leeggangers HACF, Visser L, de Vries RP: **Prevalence of transcription factors in ascomycete and basidiomycete fungi.** *BMC Genomics* 2014, **15:**214.

86. Gournas C, Prévost M, Krammer E-M, André B: **Function and Regulation of Fungal Amino Acid Transporters: Insights from Predicted Structure.** In *Yeast Membrane Transport.* Edited by Ramos J, Sychrová H, Kschischo M. Cham: Springer International Publishing; 2016: 69-106

87. Brunner K, Zeilinger S, Ciliento R, Woo SL, Lorito M, Kubicek CP, Mach RL: **Improvement of the fungal biocontrol agent Trichoderma atroviride to enhance both antagonism and induction of plant systemic disease resistance.** *Applied and environmental microbiology* 2005, **71:**3959-3965.

88. Abdel-Fattah GM, Shabana YM, Ismail AE, Rashad YM: **Trichoderma harzianum: a biocontrol agent against Bipolaris oryzae.** *Mycopathologia* 2007, **164:**81-89.

89. Adnan M, Islam W, Shabbir A, Khan KA, Ghramh HA, Huang Z, Chen HYH, Lu G-d: **Plant defense against fungal pathogens by antagonistic fungi with Trichoderma in focus.** *Microbial Pathogenesis* 2019, **129:**7-18.

90. 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.** *Bioresource Technology* 2012, **107:**517-521.

91. da Silva Delabona P, Rodrigues GN, Zubieta MP, Ramoni J, Codima CA, Lima DJ, Farinas CS, da Cruz Pradella JG, Seiboth B: **The relation between xyr1 overexpression in Trichoderma harzianum and sugarcane bagasse saccharification performance.** *Journal of Biotechnology* 2017, **246:**24-32.

92. Te Velthuis AJW, Bagowski CP: **Linking fold, function and phylogeny: a comparative genomics view on protein (domain) evolution.** *Current genomics* 2008, **9:**88-96.

93. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95:**14863-14868.

94. van Peij NN, Gielkens MM, de Vries RP, Visser J, de Graaff LH: **The transcriptional activator XlnR regulates both xylanolytic and endoglucanase gene expression in Aspergillus niger.** *Applied and environmental microbiology* 1998, **64:**3615-3619.

95. Kowalczyk JE, Benoit I, de Vries RP: **Chapter Two - Regulation of Plant Biomass Utilization in Aspergillus.** In *Advances in Applied Microbiology. Volume* 88. Edited by Sariaslani S, Gadd GM: Academic Press; 2014: 31-56

96. Sun J, Tian C, Diamond S, Glass NL: **Deciphering transcriptional regulatory mechanisms associated with hemicellulose degradation in Neurospora crassa.** *Eukaryotic cell* 2012, **11:**482-493.

97. Tanaka K: **The proteasome: overview of structure and functions.** *Proceedings of the Japan Academy Series B, Physical and biological sciences* 2009, **85:**12-36.

98. Sylvester JE, Fischel-Ghodsian N, Mougey EB, O'Brien TW: **Mitochondrial ribosomal proteins: Candidate genes for mitochondrial disease.** *Genetics in Medicine* 2004, **6:**73-80.

99. Gray JH, Owen RP, Giacomini KM: **The concentrative nucleoside transporter family, SLC28.** *Pflügers Archiv* 2004, **447:**728-734.

100. Hanas JS, Larabee JL, Hocker JR: **Zinc Finger Interactions with Metals and Other Small Molecules.** In *Zinc Finger Proteins: From Atomic Contact to Cellular Function.* Edited by Iuchi S, Kuldell N. Boston, MA: Springer US; 2005: 39-46

101. Takatsuji H: **Zinc-finger transcription factors in plants.** *Cellular and Molecular Life Sciences CMLS* 1998, **54:**582-596.

102. Sivasithamparam KG, Emilio.: **Secondary metabolism in Trichoderma and Gliocladium.** In *Trichoderma and Gliocladium. Volume* 1. Edited by Harman. CPKGE. London, UK; 1998: pp. 139-191

103. Mukherjee PK, Nautiyal CS, Mukhopadhyay AN: **Molecular Mechanisms of Biocontrol by Trichoderma spp.** In *Molecular Mechanisms of Plant and Microbe Coexistence.* Edited by Nautiyal CS, Dion P. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008: 243-262

104. Antoniêto ACC, de Paula RG, Castro LDS, Silva-Rocha R, Persinoti GF, Silva RN: **Trichoderma reesei CRE1-mediated Carbon Catabolite Repression in Response to Sophorose Through RNA Sequencing Analysis.** *Current genomics* 2016, **17:**119-131.

105. Portnoy T, Margeot A, Linke R, Atanasova L, Fekete E, Sándor E, Hartl L, Karaffa L, Druzhinina IS, Seiboth B, et al: **The CRE1 carbon catabolite repressor of the fungus Trichoderma reesei: a master regulator of carbon assimilation.** *BMC Genomics* 2011, **12:**269.

106. Martens-Uzunova ES, Schaap PJ: **Assessment of the pectin degrading enzyme network of Aspergillus niger by functional genomics.** *Fungal Genetics and Biology* 2009, **46:**S170-S179.

107. de Vries RP, Visser J: **Aspergillus enzymes involved in degradation of plant cell wall polysaccharides.** *Microbiology and molecular biology reviews : MMBR* 2001, **65:**497-522.

108. Nogueira KMV, de Paula RG, Antoniêto ACC, dos Reis TF, Carraro CB, Silva AC, Almeida F, Rechia CGV, Goldman GH, Silva RN: **Characterization of a novel sugar transporter involved in sugarcane bagasse degradation in Trichoderma reesei.** *Biotechnology for Biofuels* 2018, **11:**84.