Network analysis reveals different strategies of Trichoderma spp. associated with XYR1 and CRE1 during cellulose degradation

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], Anete Pereira de Souza[1,3*]

[1] Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, 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, Cidade Universitária Zeferino Vaz, Rua Monteiro Lobato, 255, Campinas, SP, Brazil

* Corresponding author

Email: anete@unicamp.br (APS)

Abstract

Trichoderma atroviride and *Trichoderma harzianum* are mycoparasitic fungi 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 by the transcription factors (TFs) XYR1 and CRE1. Herein, we performed a weighted correlation network analysis (WGCNA) for *T. harzianum* (IOC-3844 and CBMAI-0179) and *T. atroviride* (CBMAI-0020) to explore the genetic mechanisms of XYR1 and CRE1 under cellulose degradation conditions. In addition, we explored the evolutionary relationships among ascomycete fungi by inferring phylogenetic trees based on the protein sequences of both regulators. Transcripts encoding carbohydrate-active enzymes (CAZymes), TFs, sugar and ion transporters, kinases, phosphatases, and proteins with unknown function were coexpressed with *xyr1* or *cre1*, and several metabolic pathways were recognized with high similarity between both regulators. Hubs from these groups included transcripts not yet characterized or described as related to cellulose degradation. Furthermore, the phylogenetic analyses indicated that XYR1 and CRE1 are extensively distributed among ascomycete fungi and suggested that both regulators from *T. atroviride* are differentiated from those from *T. harzianum*. The set of transcripts coexpressed with *xyr1* and *cre1* differed by strain, suggesting that each evaluated fungus uses a specific molecular mechanism for cellulose degradation. This variation was accentuated between *T. atroviride* and *T. harzianum* strains, supporting the phylogenetic analyses of CRE1 and XYR1. Our results improve our understanding of the regulation of hydrolytic enzyme expression and

highlight the potential of *T. harzianum* for use in several biotechnological industrial applications, including bioenergy and biofuel production.

Introduction

Fungi of the genus *Trichoderma* are characterized by considerable nutritional versatility [1], which allows their employment in a wide range of biotechnological applications [2]. While fungal cellulases from *Trichoderma reesei* are the most applied enzymes in the biotechnology industry [3], the use of *Trichoderma harzianum* and *Trichoderma atroviride* was explored by examining their biocontrol capacity against plant pathogenic fungi [4,5]. Recently, the potential of enzymes from *T. harzianum* strains was demonstrated in the conversion of lignocellulosic biomass into sugars during second-generation ethanol (2G ethanol) production [6,7].

Lignocellulosic biomass is a complex recalcitrant structure that requires a consortium of carbohydrate-active enzymes (CAZymes) [8] for its complete depolymerization. The CAZymes that act on glycosidic bonds have proven to be crucial for significant biotechnological advances within sectors that involve biofuels and high-value biochemicals. In addition, CAZymes are predicted to be important components of mycoparasitism, being involved in the degradation of the cell wall of fungal prey species [9]. Based on their catalytic activities, CAZymes are divided into several classes, including glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), glycosyltransferases (GTs), and auxiliary activities (AA). In relation to their functional amino acid sequences and structural similarities, they are further classified into several families (http://www.cazy.org/).

Filamentous fungi are widely explored for the industrial production of CAZymes due to their unique ability to secrete these proteins efficiently in the presence of the plant

polysaccharide they specifically act on. The presence of the inducer results in the activation of a substrate-specific transcription factor (TF), which is required for the controlled expression of the genes encoding the CAZymes, transporters, and metabolic pathways needed to metabolize the resulting monomers [10]. In *T. reesei*, the Zn_2Cys_6 type TF xylanase regulator 1 (XYR1) is the most important activator [11]. Apart from activators, several repressor proteins involved in plant biomass degradation have been identified [10]. In the presence of easily metabolizable carbon sources, such glucose, the expression of *xyr1* and of genes encoding lignocellulose-degrading enzymes is repressed by carbon catabolite repression (CCR), which is regulated by C_2H_2 type TF carbon catabolite repressor 1 (CRE1) [12-14].

Although XYR1 orthologues are present in almost all filamentous ascomycete fungi, the molecular mechanisms triggered by this regulator depend on the species [15]. XYR1 was first described in *T. reesei* as the main activator of enzymes involved in d-xylose metabolism [11]. 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 [16]. CRE1 is the only conserved TF throughout the fungal kingdom, suggesting a conserved mechanism for CCR in fungi [17].

Plant biomass-degrading fungi are found throughout the fungal kingdom, and their different lifestyles are often well reflected in the set of genes encoding CAZymes [18]. Even related fungi with similar genome content produce highly diverse enzyme sets when grown on the same plant biomass substrate [19]. Thus, the regulation of gene expression appears to have a

dominant effect on the diversity of plant biomass-degradation strategies of fungi. One interesting *in silico* approach to exploring the regulation of lignocellulolytic enzymes is through gene coexpression networks, which have proven to be a fundamental tool for identifying simultaneously active transcripts and, consequently, inferring gene interactions in several biological contexts [20,21].

Weighted correlation network analysis (WGCNA) is a systems biology approach for the investigation of complex gene regulatory networks [22]. Based on RNA-sequencing data, WGCNA integrates the expression differences between samples into a higher-order network structure and elucidates the relationships between genes based on their correlated expression patterns [22]. To date, only a few studies have employed WGCNA method in studies on *Trichoderma* spp. [23] and on other filamentous fungi [24,25].

Here, we aimed to investigate how the TFs XYR1 and CRE1 act in *Trichoderma* spp. during cellulose degradation. 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) 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 conditions. The insights obtained will be crucial for expanding the use of *T. harzianum* as a hydrolytic enzyme producer for the development of better enzyme cocktails not only for biofuel and biochemical production but also for several industrial applications.

Materials and methods

Fungal strains, culture conditions and transcription profiling

The identity of *Trichoderma* isolates was authenticated by the sequences of their internal transcribed spacer (ITS) region and translational elongation factor 1 (*tef1*) marker genes. The culture conditions and transcription profiling of *T. harzianum* CBMAI-0179 (Th0179), *T. harzianum* IOC-3844 (Th3844), and *T. atroviride* CBMAI-0020 (Ta0020) strains are described in Additional file 1.

Phylogenetic analyses

The ITS nucleotide sequences from *Trichoderma* spp. were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/). Additionally, ITS nucleotide sequences amplified from the genomic DNA of Th3844, Th0179, and Ta0020 via PCR were kindly provided by the Brazilian Collection of Environment and Industry Microorganisms (CBMAI, Campinas, Brazil) and included in the phylogenetic analysis. The ascomycete fungus CRE1 and XYR1 protein sequences with similarities ≥ 60% and ≥ 70%, respectively, to the reference genome of *T. harzianum* T6776 were retrieved from the NCBI database. Additionally, CRE1 and XYR1 sequences were recovered from the *T. harzianum* T6776 [26] reference genome for Th3844 and Th0179 and the *T. atroviride* IMI206040 [27] genome for Ta0020 and used as references for the

TF consensus sequences for RNA-Seq read mapping. These sequences were included in the phylogenetic analyses.

Multiple sequence alignment was performed using ClustalW [28], and a phylogenetic tree was created using Molecular Evolutionary Genetics Analysis (MEGA) software v7.0 [29]. The maximum likelihood (ML) [30] method of inference was used based on a (I) Kimura two-parameter (K2P) model [31], (II) Dayhoff model with freqs (F^+), and (III) Jones-Taylor-Thornton (JTT) model with freqs (F^+) for ITS, CRE1, and XYR1, respectively. We used 1,000 bootstrap replicates [32] for each analysis. The trees were visualized and edited using FigTree 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 transcripts per million (TPM) values with the R [33] WGCNA package [22]. 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 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. Next, an adjacency matrix was obtained in which nodes correspond to transcripts and edges correspond to the strength of their connection. To obtain a dissimilarity matrix, we built a topological overlap matrix (TOM) as implemented in the package.

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. From the acquired dendrogram, the dynamicTreeCut package [34] was used to obtain the ideal number of groups and the respective categorization of the transcripts. According to the functional annotation performed by Almeida et al. (2020) [35], 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 the Python 3 programming language [36] together with the networkx library [37] to identify subgroups using the Louvain method [38] and hubs based on the degree of distribution. We used different thresholds to keep edges on the networks and recognize 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, for subgroup identification, we tested the Molecular Complex Detection (MCODE) (v1.4.2) plugin [39] using Cytoscape software v3.7.2 with default parameters.

Functional annotation of transcripts within xyr1 and cre1 groups

We conducted functional annotation of the transcripts within the *xyr1* and *cre1* groups. All identified Gene Ontology (GO) [40] categories were used to create a treemap to visualize possible correlated categories in the dataset caused by cellulose degradation using the REVIGO

tool [41]. The metabolic pathways related to the Kyoto Encyclopedia of Genes and Genomes (KEGG) [42] Orthology (KO) identified in *T. reesei* were then selected due the high number of annotations correlated with this species in the KEGG database. To obtain the pathways related to the enzymes identified as belonging to different groups, we used the Python 3 programming language [36] together with the BioPython library [43]. The automatic annotation was revised manually using the UniProt databases [44].

Results

Molecular phylogeny of *Trichoderma spp.*

The strains studied were phylogenetically classified (Additional file 1: S1 Fig). The phylogenetic tree illustrated a high phylogenetic proximity between Th3844 and Th0179, which presented a distant relationship with Ta0020. *Neurospora crassa* and *Fusarium oxysporum* were used as an outgroup.

Molecular phylogeny of CRE1 and XYR1 in ascomycota

The phylogenetic analysis of the amino acid sequences related to CRE1 and XYR1 revealed a clear diversification of these regulators inside the ascomycete fungi (Fig 1 and Additional file 2: S1 Table). The families represented in the CRE1 phylogenetic tree (Fig 1A) were not the same as those identified for XYR1 (Fig 1B).

Fig 1. Molecular Phylogenies of CRE1 and XYR1 in Ascomycota. The complete protein sequences related to CRE1 (A) and XYR1 (B) were used to infer the phylogenetic relationships of the studied *Trichoderma* species/strains within the ascomycete fungi. The phylogenetic trees were constructed using MEGA7 software and edited using the FigTree program. Th3844: *T. harzianum* IOC-3844; Th0179: *T. harzianum* CBMAI-0179; Ta0020: *T. atroviride* CBMAI-0020.

Among *Trichoderma* spp., a differentiation process for Ta0020 compared to Th3844 and Th0179 was observed for both regulators (Fig 1). In the CRE1 phylogenetic tree, close phylogenetic proximity between Th3844 and Th0179 (with bootstrap support of 51%) was observed. In contrast, Ta0020 shared closer affinity with two other *T. atroviride* strains (XP_013941427.1 and AKM12665.1), with bootstrap support of 99% (Fig 1A). In the XYR1 phylogenetic tree, Th3844, Th0179, and Th0179 were closest to *T. harzianum* CBS 226.95 (XP_024780881.1) (with bootstrap support of 96%), *T. harzianum* T6776 (KKO98630.1) (with bootstrap support of 50%), and *T. atroviride* IMI206040 (XP_013941705.1) (with bootstrap support of 88%), respectively (Fig 1B).

Construction of gene coexpression networks for *Trichoderma spp.*

By using the transcriptome data, we obtained (I) 11,050 transcripts (Th3844), (II) 11,105 transcripts (Th0179), and (III) 11,021 transcripts (Ta0020) (Additional file 3: S2 Table). Different softpower β values were chosen to fit the scale independence of the different networks

to a scale-free topology (48 for Th3844, 8 for Th0179, and 27 for Ta0020). 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 4: S3 Table).

Coexpressed transcripts with XYR1 and CRE1

Transcripts with expression patterns correlated with XYR1 and CRE1 were identified in the Th3844, Th0179, and Ta0020 coexpression networks and grouped together (described in Additional file 5: S4 Table). Among the strains, Ta0020 presented the highest number of transcripts coexpressed with *cre1* and the lowest number coexpressed with *xyr1*; the other strains showed the opposite profile.

Each group containing the *xyr1* and *cre1* transcripts was partitioned, and subgroups were formed for all strains. The numbers of subgroups identified through the MCODE algorithm are presented in Additional file 6: S5 Table. The transcripts encoding CRE1 and XYR1 were only found for Th3844 (*cre1* and *xyr1*) and Th0179 (*cre1*). Then, we employed the Louvain algorithm to detect the more correlated niches in the *cre1* and *xyr1* groups (Fig 2 and Additional file 7: S6 Table).

Fig 2. Subgroups Identified in the *cre1* and *xyr1* Groups among *Trichoderma* spp. The selected groups were partitioned according to the Louvain algorithm, and subgroups of Th3844 (A), Th0179 (B) and Ta0020 (C) related to *cre1* and for Th3844 (D), Th0179 (E) and Ta0020 (F)

related to *xyr1* were formed. Different colors represent different subgroups within a group. Across groups, the same 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 inside a group, an equal color was attributed to the transcripts that constituted the respective subgroup, which was denoted according to the strain (Additional file 1: S7 Table). The *cre1* transcript was in the **1.0**, **2.1**, and **3.5** subgroups for Th3844, Th0179, and Ta0020, respectively. 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. In addition, we observed the presence of disconnected elements in the *cre1* subgroups formed for Th0179 and Ta0020. This same pattern was observed in the *xyr1* group for Ta0020. 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 only a few transcripts.

Hub transcripts with expression patterns similar to XYR1 and CRE1

Transcripts with a high number of coexpressed neighbors, defined here as hubs, were sought in *cre1* and *xyr1* groups and are available in Additional file 8: S8 Table. The size of the correlation coefficient threshold variate was used to identify highly connected transcripts within a group (0.5 represents moderate correlation and 0.9 represents relatively high correlation) [45].

Functional annotation of transcripts within xyr1 and cre1 groups

We performed enrichment analysis using the GO categories with the transcripts within each *cre1* and *xyr1* group for Th3844, Th0179, and Ta0020 (Additional file 1: S2-S4 Figs, respectively). For both groups, the number of exclusive enriched terms for each strain was larger than the number of shared terms, which indicated the high degree of differences among transcripts. In relation to metabolic pathways, we selected KO functional annotation of the transcripts encoding enzymes and unknown proteins with enzymatic activity enriched in the *cre1* and *xyr1* groups. For the 14 pathway classes observed for each TF, 13 were shared between the two regulators (Fig 3). However, the number of identified enzymes for a given pathway class was different for each strain.

Fig 3. KO Functional Classification of the Transcripts Identified in the *cre1* and *xyr1* Groups. The sequences related to enzymes 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.

Differentially expressed transcripts from the xyr1 and cre1 groups

Different quantities of differentially expressed transcripts were found for Ta0020 (78) and Th0179 (6) (described in Additional file 9: S9 Table and Additional file 1: S10 Table). For Ta0020, most of the upregulated transcripts under cellulose growth (70) were associated with the *cre1* group, including *cre1*, whereas *xyr1* demonstrated a potentially significant expression level

under glucose growth. For Th3844, *cre1* showed a significant modulation in expression between the two carbon sources, with a high expression level under glucose growth. Even with the reduced differentiation observed (-1.13), this phenomenon was considered by other studies [46,47]. For Th0179, *xyr1* and *cre1* revealed no significant modulation in transcript expression between growth under cellulose or glucose. However, *xyr1* had a higher transcript expression level under glucose compared to *cre1*, which had a higher transcript expression level under cellulose growth conditions. For Th3844, similar results were observed for *xyr1*, although a higher transcript expression level was observed under glucose.

CAZyme, TF and transporter distributions among the 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). Transcripts encoding TFs, transporters, and CAZymes in the *cre1* and *xyr1* subgroups are described in Additional file 7: S6 Table.

Fig 4. Comparisons of TFs, Transporters, and CAZymes among Strains and Groups. Number of transcripts encoding TFs, transporters, and CAZymes for Ta0020, Th0179 and Th3844 distributed among the *cre1* (A) and *xyr1* (B) groups. Transcripts encoding the proteins associated with *xyr1* were more widely distributed among all strains than were those associated with *cre1*. TFs: transcription factors; CAZymes: carbohydrate-active enzymes. Ta0020: *T.*

atroviride CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844.

CAZymes

GHs were identified in all of the evaluated strains with different numbers (Fig 5). For the *cre1* group, while Ta0020 presented the highest number of classified transcripts from the GH family (Fig 5A), Th3844 presented only one transcript encoding GH28. For the *xyr1* group, this last strain presented a higher number of transcripts from the GH family than did Th0179 (GH16 and GH76) and Ta0020 (GH75) (Fig 5B). CEs were associated with the *cre1* transcript in Ta0020 and Th0179 (Fig 5C and D), while polysaccharide lyase (PL) was coexpressed with the *xyr1* transcript in only Th3844 (Fig 5E). Glycosyltransferase (GT) was coexpressed with the *cre1* transcript in only Th0179 (Fig 5D) and with the *xyr1* transcript in Th3844 (Fig 5E) and Ta0020 (GT22).

Fig 5. Distribution of CAZyme Families within the *cre1* and *xyr1* Groups in *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.

TFs

Zn_2Cys_6 was the most abundant coexpressed transcript for the strains that showed transcripts encoding TFs in the *cre1* and *xyr1* groups (Fig 6A and B). In the *cre1* group for Ta0020, the number of transcripts encoding Zn_2Cys_6 was higher than that observed for Th0179. For the *xyr1* group, both Th3844 and Th0179 produced several more transcripts encoding Zn_2Cys_6 than Ta0020. We found transcripts encoding a fungal-specific TF domain for Ta0020 and Th3844 in the *cre1* and *xyr1* groups, respectively. Transcripts encoding C_2H_2 were identified in the *cre1* group for Ta0020 and in the *xyr1* group for *T. harzianum* strains. Furthermore, we found a transcript encoding the SteA regulator only in the *xyr1* group for Th0179.

Fig 6. Distribution of TFs and transporters within *cre1* and *xyr1* groups in *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. MFS: major facilitator superfamily; ABC: ATP-binding cassette; Th3844: *T. harzianum* IOC-3844; Th0179: *T. harzianum* CBMAI-0179; Ta0020: *T. atroviride* CBMAI-0020.

Transporters

Transcripts encoding ion transporters were present in the *cre1* and *xyr1* groups for Ta0020 and Th0179 (Fig 6C and D). In the *cre1* group, major facilitator superfamily (MFS) permeases were the most abundant proteins among the transporters for Ta0020 (Fig 6C). In the *xyr1* group, only Th0179 showed transcripts encoding this type of protein transporter (Fig 6D). Th3844 showed transcripts encoding ATP-binding cassette (ABC) transporters and amino acid

transporters with a similar expression pattern as *xyr1* (Fig 6D). In addition, all strains presented transcripts encoding amino acid transporters coexpressed with *xyr1* (Fig 6D). For the *cre1* group, only Th0179 showed transcripts encoding this type of transport protein (Fig 6C).

Transcripts related to protein phosphorylation coexpressed with *cre1* and *xyr1*

Given the significance of kinases and phosphatases in diverse fungi physiological functions, including hemicellulase and cellulase gene expression, we selected transcripts encoding these types of proteins coexpressed with *cre1* and *xyr1*. In the *cre1* group, Th0179 and Ta0020 presented transcripts encoding hypothetical proteins with kinase activity. Ta0020 also showed transcripts encoding several types of kinases proteins, as did Th3844. In the *xyr1* group, Th3844 presented only transcripts encoding hypothetical proteins with kinase activity, whereas Ta0020 and Th0179 exhibited transcripts encoding a number of kinases proteins. In the *cre1* group and *xyr1* group, Ta0020 and Th0179, respectively, exhibited a higher number of coexpressed transcripts encoding proteins with kinase activity than did the other strains.

All strains showed transcripts encoding phosphatases coexpressed with *cre1*. Ta0020 exhibited a higher number than the other strains. In the *xyr1* group, Th0179 and Th3844 showed transcripts encoding various categories of phosphatases. Th3844 also presented transcripts encoding hypothetical proteins, and Ta0020 exhibited just one hypothetical protein. The

transcripts encoding kinases and phosphatases in the *cre1* and *xyr1* groups and subgroups are described in Additional file 10: S11 Table and Additional file 7: S6 Table.

Discussion

Genetic and enzymatic diversity within the genus *Trichoderma*

Fungi have distinct regulatory systems that control the expression and secretion of genes encoding enzymes that degrade plant cell walls [48]. Even closely related species produce highly diverse enzyme sets when grown on the same plant biomass substrate [19]. The great variability found within the genus *Trichoderma* suggests a high level of adaptation of individual species [49], which results in contrasting carbon metabolism profiles [50]. Thus, a comparison of genome-wide expression at the transcript level in *Trichoderma* spp. could lead to the identification of the genetic mechanisms underlying biomass degradation. Herein, we suggest that the three evaluated strains present different sets of transcripts that are coexpressed with the key regulators of biomass degradation, which could contribute to the observed contrasting enzymatic performances during cellulose deconstruction.

The regulation of genes involved in cellulose degradation in *Trichoderma spp.* and evolutionary insights

Studies have reported the feasibility of enhancing cellulase production by genetically modifying the expression of the main regulators involved in biomass degradation. Interestingly, the effect of *cre1* deletion on the gene expression profile of its targets varies among species. In the *T. reesei* RUT-C30 wild-type strain, full *cre1* deletion led to pleiotropic effects and strong growth impairment, whereas truncation of *cre1*-96 enhanced cellulolytic activity without causing growth deficiencies [51]. In contrast, in the *T. harzianum* P49P11 Δ*cre1* mutant, the expression levels of CAZyme genes and *xyr1* were significantly higher than those in the wild-type strain [52]. However, the absence of *cre1* has been found to have negative effects on the production of some CAZyme genes, such as *gh61*, *bgl1*, and *xyn2*, suggesting that CRE1 may act positively on genes essential for biomass degradation [52]. Yet using *T. harzianum* P49P11 as a fungal model, Delabona et al. (2017) [53], constructed strains with *xyr1* overexpression. Although the genetic modification of *xyr1* increased the levels of reducing sugars released in a shorter time during saccharification, CRE1 was upregulated [53].

Although the importance of XYR1 and CRE1 is evident, the transcriptional regulation involved in the expression of CAZyme-encoding genes in *T. harzianum* strains remains poorly explored relative to that in the major producer of cellulases, *T. reesei*. Almeida et al. (2020) [35] reported different types of enzymatic profiles among *Trichoderma* species, with Ta0020 having a lower cellulase activity during growth on cellulose than Th3844 and Th0179. Using the same transcriptional data, we investigated and compared the genetic mechanisms involved in XYR1- and CRE1-mediated gene regulation for these strains by modeling gene coexpression networks.

Due to the absence of an available reference genome for the Th3844, Th0179, and Ta0020 strains, *T. harzianum* T6776 and *T. atroviride* IMI 206040 were used as references for transcriptome assembly [35]. The variation regarding the number of mapped reads on the different references was 0.72% for Th0179 and 1.19% for Th3844, supporting the use of the *T. harzianum* T6776 genome as reference [54-56]. Close phylogenetic proximity was observed among the T6776, CBS 226.95, and TR257 strains of *T. harzianum*, which exhibit similar genome sizes as well as similar numbers of orthologues and paralogous genes [49].

The usage of molecular phylogenetic approaches is critical to advance the understanding of protein diversity and function from an evolutionary perspective [57]. The results of a previous study [58] suggested that some genes encoding regulatory proteins, such as *xyr1*, evolved by vertical gene transfer. Herein, we study the phylogenetic relationships of XYR1 and CRE1, including those from our studied strains. For both regulators, we noted a high diversification inside Ascomycota. Additionally, a differentiation process was observed for Ta0020 compared to Th3844 and Th0179, which supports the molecular phylogeny of the species based on the ITS region.

WGCNA as a bioinformatic tool to explore gene regulation

Recently, WGCNA was used to explore the molecular mechanisms underlying biomass degradation by *Penicillium oxalicum* [25]. In addition, in a large-scale proteomics study, WGCNA was applied to investigate and compare the enzymatic responses of the filamentous

fungi *Aspergillus terreus*, *T. reesei*, *Myceliophthora thermophila*, *N. crassa*, and *Phanerochaete chrysosporium* grown in several carbon sources [24]. Within the genus *Trichoderma*, WGCNA was employed to identify the set of transcripts used by *T. reesei* during sugarcane bagasse breakdown [23]. Such studies demonstrate that the WGCNA methodology can clarify the molecular mechanisms that underlie biological processes and reveal how changes in gene interactions might lead to an altered phenotype [20]. By employing WGCNA in the present study, we were able to cluster transcripts functionally related to CRE1 and XYR1 and identify the hubs in the groups of interest. These groups can serve as resources for generating new testable hypotheses and prediction, including the prediction of gene function and importance in a biomass degradation context.

The metabolic diversity of fungi, particularly with regards to their secondary metabolism, which involves ecologically important metabolites not essential to cellular life, is key to explaining their evolutionary success [59]. XYR1 and CRE1 were associated with distinct metabolic processes in Ta0020 and Th0179, which strongly suggests that these two TFs have developed different regulatory functions on these strains. In addition, the exclusive pathway classes for Ta0020 were not directly related to the biomass degradation process. For example, the metabolism of terpenoids and polyketides, two secondary metabolite compounds involved in the antibiosis process during mycoparasitic activity [60], was identified.

The predictive potential of hubs involved in biomass degradation by *Trichoderma* spp. remains to be thoroughly investigated. Proteins related to the proteasome, oxidative

phosphorylation, and nucleoside transport were found to be hubs in Th3844, Th0179, and Ta0020, respectively, in the *cre1* groups. Transcripts encoding uncharacterized proteins were identified as potential hub nodes in the *xyr1* groups. The functional data on these hubs allow their accurate placement in the overall network of fungal plant biomass degradation.

The functional composition of *xyr1* and *cre1* groups in different strains

We also investigated the expression profiles of the transcripts within the *cre1* and *xyr1* groups, including *xyr1* and *cre1*, under cellulose and glucose growth. Castro et al. (2016) [47] noted that in *T. reesei*, the expression level of *xyr1* was minimal in the presence of glucose; this phenomenon was not observed for Ta0020 in the present study. In addition, in Ta0020, *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 [10]. Conversely, for Th3844, the *cre1* expression level was higher in the presence of glucose than in the presence of cellulose.

Genes can evolve for divergent functions in different species, even when exhibiting high similarity in sequence or protein domains [61]. For *T. reesei*, the expressions of CAZymes, TFs, and transporters were affected by XYR1 in the context of plant biomass degradation [47], whereas for *T. atroviride*, this TF was identified as a critical factor during the plant-fungus recognition processes and responsible for mycoparasitism activity [16]. Although transcripts

encoding these proteins coexpressed with *xyr1* were present in Ta0020, they may induce different mechanisms in Th3844 and Th0179, as reported for ascomycete fungi [15].

Among CAZymes, GHs are the most abundant enzymes for the breakdown of glycosidic bonds between carbohydrate molecules [62]. Several GHs were also identified during the mycoparasitism process, and their activities may be related to fungal cell wall degradation in mycoparasite species of *Trichoderma* [63]. Our results showed that *T. atroviride* has an increased number of GHs coexpressed with *cre1* that are related to mycoparasitism and only a few associated with cellulose and hemicellulose degradation. In the *T. harzianum* strains, the main CAZyme classes were not found in the *cre1* and *xyr1* groups.

In *T. reesei* grown on steam-exploded sugarcane bagasse, Borin et al. (2018) [23] found transcripts encoding GHs that were coexpressed with some of the TFs involved in biomass degradation, e.g., XYR1. However, the transcriptome dataset used for this study was based on a different carbon source from that employed by Borin et al. (2018) [23]. Previous studies suggest that the gene expression and protein secretion of CAZymes is substrate-specific [24,64,65]. Accordingly, in the present study, the transcripts encoding hydrolytic enzymes could be sorted into other groups formed within the coexpression networks, mainly because their expression patterns were different from those observed for the transcripts encoding XYR1 and CRE1. Furthermore, biomass breakdown is a complex process that involves several signaling pathways, not only those related to the regulation of genes encoding CAZymes. A detailed examination of the regulation of cellulolytic enzymes in *T. reesei* evidenced significant functions of the Ca^{2+}

sensing system and the regulator CRZ1 in the control of XYR1 and CRE1 [66]. These results support our findings that the transcripts predicted to be coexpressed with the studied TFs were not comprised only of transcripts encoding CAZymes, since other factors influence the complex regulation of hydrolytic enzyme genes.

During plant biomass degradation, fungi secrete extracellular enzymes to decompose the polysaccharides to small molecules, which are then imported into cells through transporters and used to support fungus development. Thus, transporters have an essential role in the utilization of lignocellulose by fungi [67]. Several important fungal TFs involved in the degradation of plant biomass polysaccharides have been reported to regulate the expression of transporters genes. In *T. reesei*, MFS, amino acids, and ABC transporters have been found to be modulated by XYR1 and CRE1 [46,47]. Our results suggest that the expression of transporters and TFs during cellulose degradation was affected by XYR1 in *T. harzianum*, as demonstrated for *T. reesei* [47]. In contrast, the influence of this TF was less evident in Ta0020, suggesting that the two mycoparasitic *Trichoderma* species have developed different mechanisms for sensing molecules in the environment.

At the molecular level, the ability of fungi to adapt to changing environments involves a complex network, which is responsible for integrating all signals and promoting a suitable gene response to produce a desirable reaction to the environmental conditions [68]. In *T. reesei*, different signaling pathways have been described to be involved in fungal development and environment sensing. Cellulase gene expression can be regulated by the dynamics of protein

phosphorylation and dephosphorylation, involving protein kinases and phosphatases [68], respectively. In filamentous fungi, phosphorylation is a prerequisite for CRE1 activity [69]. Furthermore, a relationship between the regulation of enzyme gene expression and secondary metabolism has been found in *T. reesei* [70,71]. Similar to previous reports, our results revealed the coexpression of various kinases and phosphatases with *xyr1* and *cre1* in all evaluated strains. Adequate environmental perception is a crucial step for the regulation of gene expression. Thus, our results reinforce the importance of studying the effects of intracellular signaling pathways on cellulase expression among *Trichoderma* spp.

Differences in enzyme production among related fungi in the genus *Trichoderma* have been reported [35,55] and raise questions about the stability of fungal genomes and the mechanisms that underlie this diversity. Herein, we showed that different transcript profiles are coexpressed with XYR1 and CRE1 during cellulose degradation in different fungi of the genus *Trichoderma*. This variation was accentuated between *T. atroviride* and *T. harzianum* strains, as supported by the phylogenetic analyses of CRE1 and XYR1, wherein the two regulators from Ta0020 were differentiated from those from Th3844 and Th0179. Addressing these questions through the investigation of the genetic regulatory mechanisms involved in biomass degradation is important for enhancing both our basic understanding and our application of fungal abilities.

Conclusions

Our findings suggest that XYR1 and CRE1 trigger different pathways in *Trichoderma* spp. during cellulose degradation and glucose metabolization — for both regulators, inter-and intraspecies (i.e., among *T. harzianum* strains) variation, was observed. The set of transcripts coexpressed with XYR1 and CRE1 was not limited to CAZymes and other proteins related to biomass degradation, reinforcing the importance of studying signal pathways in the recognition of diverse carbon sources. We expect that our results will contribute to a better understanding of fungal biodiversity, especially with regard to the genetic mechanisms involved in hydrolytic enzyme regulation in *Trichoderma* species. This knowledge will be indispensable for conceiving strategies for genetic manipulation and important for expanding the use of *T. harzianum* as an enzyme producer in biotechnological industrial applications, including bioenergy and biofuel applications.

Acknowledgments

We are grateful to the Brazilian Biorenewables National Laboratory (LNBR), Campinas – SP, for conducting the fermentation experiments; the Center of Molecular Biology and Genetic Engineering (CBMEG) at the University of Campinas, SP, for use of the center and laboratory space; and the São Paulo Research Foundation (FAPESP), the Coordination of Improvement of Higher Education Personnel (CAPES, Computational Biology Program), and the Brazilian National Council for Technological and Scientific Development (CNPq) for supporting the project and researchers.

References

1. 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: 749-759.
2. Kidwai MK, Nehra M. Biotechnological applications of *Trichoderma species* for environmental and food security. In: Gahlawat SK, Salar RK, Siwach P, Duhan JS, Kumar S, Kaur P, editors. Plant biotechnology: recent advancements and developments. Singapore: Springer; 2017. pp. 125-156.
3. 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.
4. Medeiros HA, Filho JVA, Freitas LG, Castillo P, Rubio MB, Hermosa R, et al. Tomato progeny inherit resistance to the nematode *Meloidogyne javanica* linked to plant growth induced by the biocontrol fungus *Trichoderma atroviride*. Sci Rep. 2017;7: 40216.
5. Saravanakumar K, Li Y, Yu C, Wang QQ, Wang M, Sun J, et al. Effect of *Trichoderma harzianum* on maize rhizosphere microbiome and biocontrol of Fusarium Stalk rot. Sci Rep. 2017;7: 1771.
6. Delabona PDS, Codima CA, Ramoni J, Zubieta MP, de Araujo BM, Farinas CS, et al. The impact of putative methyltransferase overexpression on the *Trichoderma harzianum* cellulolytic system for biomass conversion. Bioresour Technol. 2020;313: 123616.

7. Zhang Y, Yang J, Luo L, Wang E, Wang R, Liu L, et al. Low-cost cellulase-hemicellulase mixture secreted by *Trichoderma harzianum* EM0925 with complete saccharification efficacy of lignocellulose. Int J Mol Sci. 2020;21: 371.
8. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42: D490-D495.
9. Liu R, Wang Y, Li P, Sun L, Jiang J, Fan X, et al. Genome assembly and transcriptome analysis of the fungus *Coniella diplodiella* during infection on Grapevine (Vitis vinifera L.). Front Microbiol. 2020;11: 599150.
10. 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.
11. Stricker AR, Grosstessner-Hain K, Wurleitner E, Mach RL. Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. Eukaryot Cell. 2006;5: 2128-2137.
12. Alazi E, Ram AFJ. Modulating transcriptional regulation of plant biomass degrading enzyme networks for rational design of industrial fungal strains. Front Bioeng Biotechnol. 2018;6: 133.
13. Mach-Aigner AR, Pucher ME, Steiger MG, Bauer GE, Preis SJ, Mach RL. Transcriptional regulation of xyr1, encoding the main regulator of the xylanolytic and cellulolytic enzyme system in *Hypocrea jecorina*. Appl Environ Microbiol. 2008;74: 6554-6562.

14. Strauss J, Mach RL, Zeilinger S, Hartler G, Stöffler G, Wolschek M, et al. Crel, the carbon catabolite repressor protein from *Trichoderma reesei*. FEBS Lett. 1995;376: 103-107.

15. Klaubauf S, Narang HM, Post H, Zhou M, Brunner K, Mach-Aigner AR, et al. Similar is not the same: differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in filamentous fungi. Fungal Genet Biol. 2014;72: 73-81.

16. Reithner B, Mach-Aigner AR, Herrera-Estrella A, Mach RL. *Trichoderma* atroviride transcriptional regulator Xyr1 supports the induction of systemic resistance in plants. Appl Environ Microbiol. 2014;80: 5274-5281.

17. Adnan M, Zheng W, Islam W, Arif M, Abubakar YS, Wang Z, et al. Carbon catabolite repression in filamentous fungi. Int J Mol Sci. 2017;19: 48.

18. Makela MR, Donofrio N, de Vries RP. Plant biomass degradation by fungi. Fungal Genet Biol. 2014;72: 2-9.

19. Benoit I, Culleton H, Zhou M, DiFalco M, Aguilar-Osorio G, Battaglia E, et al. Closely related fungi employ diverse enzymatic strategies to degrade plant biomass. Biotechnol Biofuels. 2015;8: 107.

20. Gysi DM, Nowick K. Construction, comparison and evolution of networks in life sciences and other disciplines. J R Soc Interface. 2020;17: 20190610.

21. Pirim H. Construction of gene networks using expression profiles. In: Purohit HJ, Kalia VC, More RP, editors. Soft computing for biological systems. Singapore: Springer Singapore; 2018. pp. 67-89.
22. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9: 559.
23. Borin GP, Carazzolle MF, Dos Santos RAC, Riano-Pachon DM, Oliveira JVC. Gene Co-expression network reveals potential new genes related to sugarcane bagasse degradation in *Trichoderma reesei* RUT-30. Front Bioeng Biotechnol. 2018;6: 151.
24. Arntzen MO, Bengtsson O, Varnai A, Delogu F, Mathiesen G, Eijsink VGH. Quantitative comparison of the biomass-degrading enzyme repertoires of five filamentous fungi. Sci Rep. 2020;10: 20267.
25. Li CX, Zhao S, Luo XM, Feng JX. Weighted gene co-expression network analysis identifies critical genes for the production of cellulase and xylanase in *Penicillium oxalicum*. Front Microbiol. 2020;11: 520.
26. Baroncelli R, Piaggeschi G, Fiorini L, Bertolini E, Zapparata A, Pe ME, et al. Draft whole-genome sequence of the biocontrol agent *Trichoderma harzianum* T6776. Genome Announc. 2015;3: e00647-15.
27. 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.

28. 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 Res. 1994;22: 4673-4680.
29. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33: 1870-1874.
30. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992;8: 275-282.
31. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16: 111-120.
32. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39: 783-791.
33. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
34. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. Bioinformatics. 2008;24: 719-720.
35. Almeida DA, Horta MAC, Filho JAF, Murad NF, de Souza AP. The synergistic actions of hydrolytic genes in coexpression networks reveal the potential of *Trichoderma harzianum* for cellulose degradation. bioRxiv 20200114906529. 2020.
36. Sanner MF. Python: a programming language for software integration and development. J Mol Graph Model. 1999;17: 57-61.

37. Hagberg A, Swart P, Chult DS. Exploring network structure, dynamics, and function using networkx. In: Varoquaux G, Vaught T, Millman J, editors. Proceedings of the 7th python in science conference (SciPy 2008). Los Alamos, US: LANL; 2008. pp. 11–16.
38. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech: Theory Exp. 2008;2008: P10008.
39. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform. 2003;4: 2.
40. 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: 25-29.
41. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6: e21800.
42. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28: 27-30.
43. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25: 1422-1423.
44. UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47: D506-D515.

45. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24: 69-71.

46. Antonieto AC, Castro LDS, 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 Genet Biol. 2014;73: 93-103.

47. Castro LDS, de Paula RG, Antonieto AC, Persinoti GF, Silva-Rocha R, Silva RN. Understanding the role of the master regulator XYR1 in *Trichoderma reesei* by global transcriptional analysis. Front Microbiol. 2016;7: 175.

48. de Vries RP, Makela MR. Genomic and postgenomic diversity of fungal plant biomass degradation approaches. Trends Microbiol. 2020;28: 487-499.

49. Kubicek CP, Steindorff AS, Chenthamara K, Manganiello G, Henrissat B, Zhang J, et al. Evolution and comparative genomics of the most common *Trichoderma species*. BMC Genom. 2019;20: 485.

50. Wang C, Zhuang WY. Carbon metabolic profiling of *Trichoderma* strains provides insight into potential ecological niches. Mycologia. 2020;112: 213-223.

51. Mello-de-Sousa TM, Gorsche R, Rassinger A, Pocas-Fonseca MJ, Mach RL, Mach-Aigner AR. A truncated form of the Carbon catabolite repressor 1 increases cellulase production in *Trichoderma reesei*. Biotechnol Biofuels. 2014;7: 129.

52. da Silva Delabona P, Lima DJ, Codima CA, Ramoni J, Gelain L, de Melo VS, et al. Replacement of the carbon catabolite regulator (cre1) and fed-batch cultivation as

strategies to enhance cellulase production in *Trichoderma harzianum*. Bioresource Technology Reports. 2021;13: 100634.

53. Delabona PDS, Rodrigues GN, Zubieta MP, Ramoni J, Codima CA, Lima DJ, et al. The relation between xyr1 overexpression in *Trichoderma harzianum* and sugarcane bagasse saccharification performance. J Biotechnol. 2017;246: 24-32.
54. 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 Genom. 2017;18: 779.
55. 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.
56. Santos CA, Zanphorlin LM, Crucello A, Tonoli CCC, Ruller R, Horta MAC, et al. Crystal structure and biochemical characterization of the recombinant ThBgl, a GH1 beta-glucosidase overexpressed in *Trichoderma harzianum* under biomass degradation conditions. Biotechnol Biofuels. 2016;9: 71.
57. Gabaldon T. Evolution of proteins and proteomes: a phylogenetics approach. Evol Bioinform Online. 2007;1: 51-61.
58. Druzhinina IS, Chenthamara K, Zhang J, Atanasova L, Yang D, Miao Y, et al. Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus Trichoderma from its plant-associated hosts. PLoS Genet. 2018;14: e1007322.

59. Naranjo-Ortiz MA, Gabaldon T. Fungal evolution: cellular, genomic and metabolic complexity. Biol Rev Camb Philos Soc. 2020;95: 1198-1232.

60. Mukherjee PK, Nautiyal CS, Mukhopadhyay AN. Molecular mechanisms of biocontrol by *Trichoderma spp*. In: Nautiyal CS, Dion P, editors. Molecular mechanisms of plant and microbe coexistence. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. pp. 243-262.

61. Gerlt JA, Babbitt PC. Can sequence determine function? Genome Biol. 2000;1: REVIEWS0005.

62. 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 Res. 2009;37: D233-D238.

63. Gruber S, Seidl-Seiboth V. Self versus non-self: fungal cell wall degradation in Trichoderma. Microbiology (Reading). 2012;158: 26-34.

64. Huttner S, Nguyen TT, Granchi Z, Chin AWT, Ahren D, Larsbrink J, et al. Combined genome and transcriptome sequencing to investigate the plant cell wall degrading enzyme system in the thermophilic fungus *Malbranchea cinnamomea*. Biotechnol Biofuels. 2017;10: 265.

65. Pullan ST, Daly P, Delmas S, Ibbett R, Kokolski M, Neiteler A, et al. RNA-sequencing reveals the complexities of the transcriptional response to lignocellulosic biofuel substrates in *Aspergillus niger*. Fungal Biol Biotechnol. 2014;1: 3.

66. Martins-Santana L, Paula RG, Silva AG, Lopes DCB, Silva RDN, Silva-Rocha R. CRZ1 regulator and calcium cooperatively modulate holocellulases gene expression in *Trichoderma reesei* QM6a. Genet Mol Biol. 2020;43: e20190244.

67. Nogueira KMV, de Paula RG, Antonieto ACC, Dos Reis TF, Carraro CB, Silva AC, et al. Characterization of a novel sugar transporter involved in sugarcane bagasse degradation in *Trichoderma reesei*. Biotechnol Biofuels. 2018;11: 84.

68. Schmoll M, Dattenbock C, Carreras-Villasenor N, Mendoza-Mendoza A, Tisch D, Aleman MI, et al. The genomes of three uneven siblings: footprints of the lifestyles of three *Trichoderma species*. Microbiol Mol Biol Rev. 2016;80: 205-327.

69. de Assis LJ, Silva LP, Bayram O, Dowling P, Kniemeyer O, Kruger T, et al. Carbon catabolite repression in filamentous fungi is regulated by phosphorylation of the transcription factor CreA. mBio. 2021;12: e03146-20.

70. Beier S, Hinterdobler W, Monroy AA, Bazafkan H, Schmoll M. The kinase USK1 regulates cellulase gene expression and secondary metabolite biosynthesis in *Trichoderma reesei*. Front Microbiol. 2020;11: 974.

71. Monroy AA, Stappler E, Schuster A, Sulyok M, Schmoll M. A CRE1- regulated cluster is responsible for light dependent production of dihydrotrichotetronin in *Trichoderma reesei*. PLoS One. 2017;12: e0182530.

Supporting information

S1 Table. Descriptions of the Ascomycota Species used for the Phylogenetic Analysis of the TFs CRE1 and XYR1.

S2 Table. Filtered Transcripts used to Model the Gene Coexpression Networks among *Trichoderma* spp.

S3 Table. Transcripts Identified in the Cluster Analysis of Gene Coexpression Networks of *Trichoderma* spp.

S4 Table. Transcripts Identified as Coexpressed with *cre1* and *xyr1* in the Gene Coexpression Networks of *Trichoderma* spp.

S5 Table. Subgroups Identified in the *cre1* and *xyr1* Groups using the MCODE Methodology for All Evaluated Strains.

S6 Table. Subgroups Identified in the *cre1* and *xyr1* Groups using the Louvain Algorithm for All Evaluated Strains.

S7 Table. Number of Transcripts Coexpressed in Each Subgroup of the *cre1* and *xyr1* Groups in *Trichoderma* spp., Distinguished by Color.

S8 Table. Hubs identified within the *cre1* and *xyr1* Groups of *Trichoderma* spp.

S9 Table. Differentially Expressed Transcripts Established in the *cre1* and *xyr1* Groups for All Evaluated Strains.

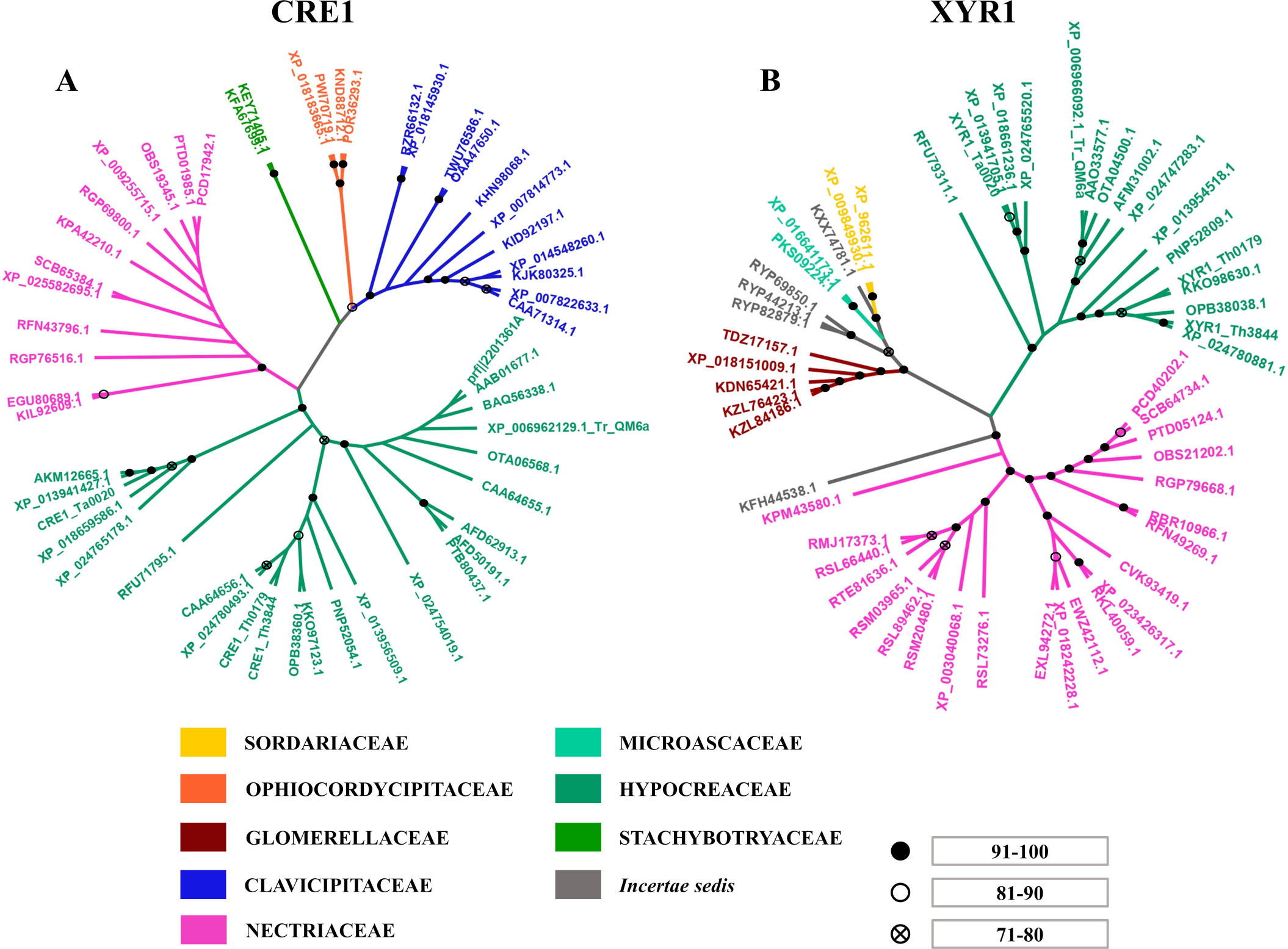
S10 Table. Expression (in TPM) of *cre1* and *xyr1* in *Trichoderma* spp. under Cellulose and Glucose Growth Conditions.

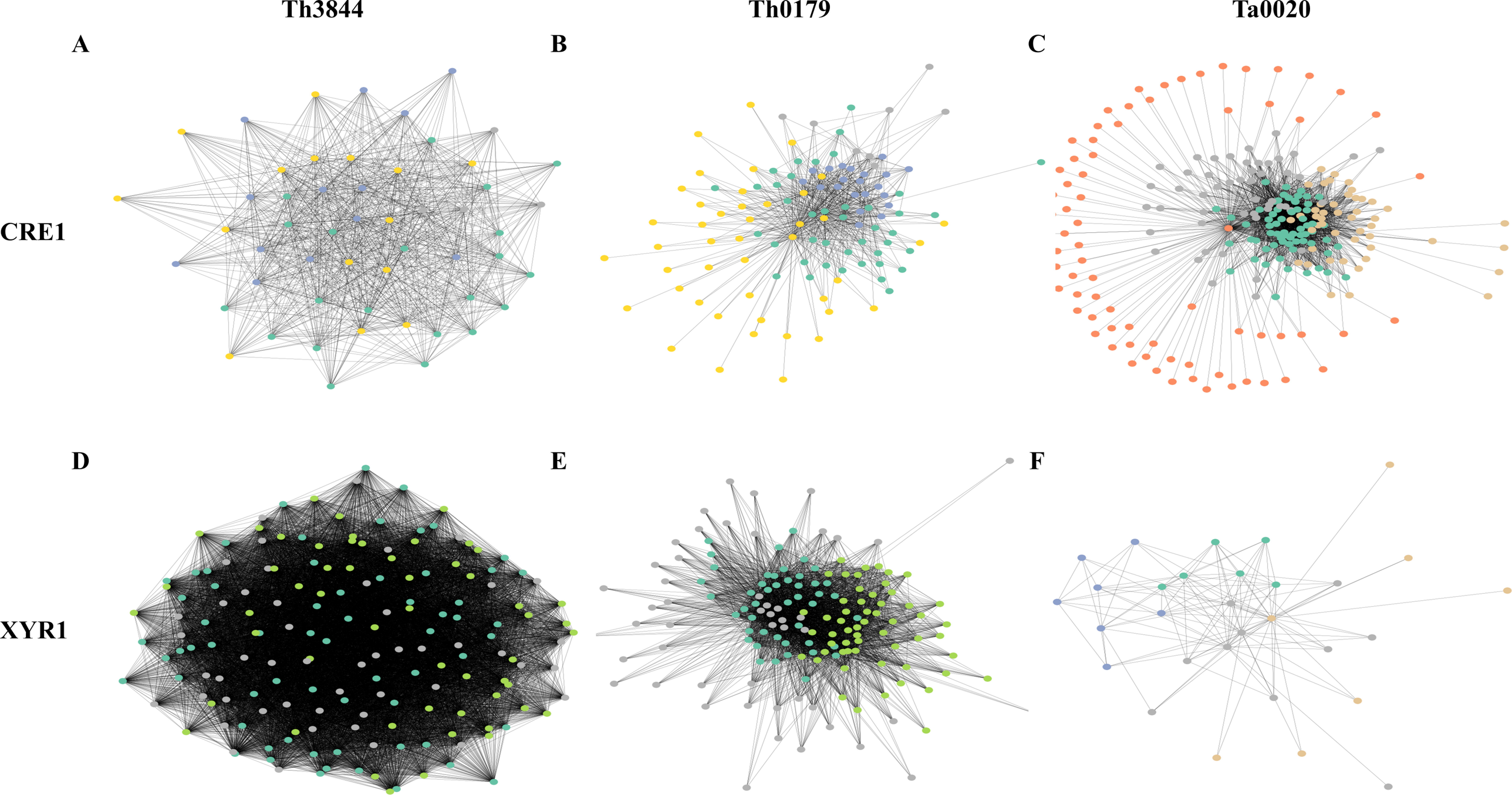
S1 Fig. Phylogenetic Relationships of *Trichoderma* spp. Inferred by Analysis of ITS Sequences. The sequences corresponding to the ITS region were used to analyze and compare the phylogenetic relationships of the studied *Trichoderma* species/strains. The ITS sequences for Th3844, Th0179, and Ta0020 were amplified from genomic DNA. The other ITS sequences were derived from the NCBI database. Th3844: *T. harzianum* IOC-3844; Th0179: *T. harzianum* CBMAI-0179; Ta0020: *T. atroviride* CBMAI-0020.

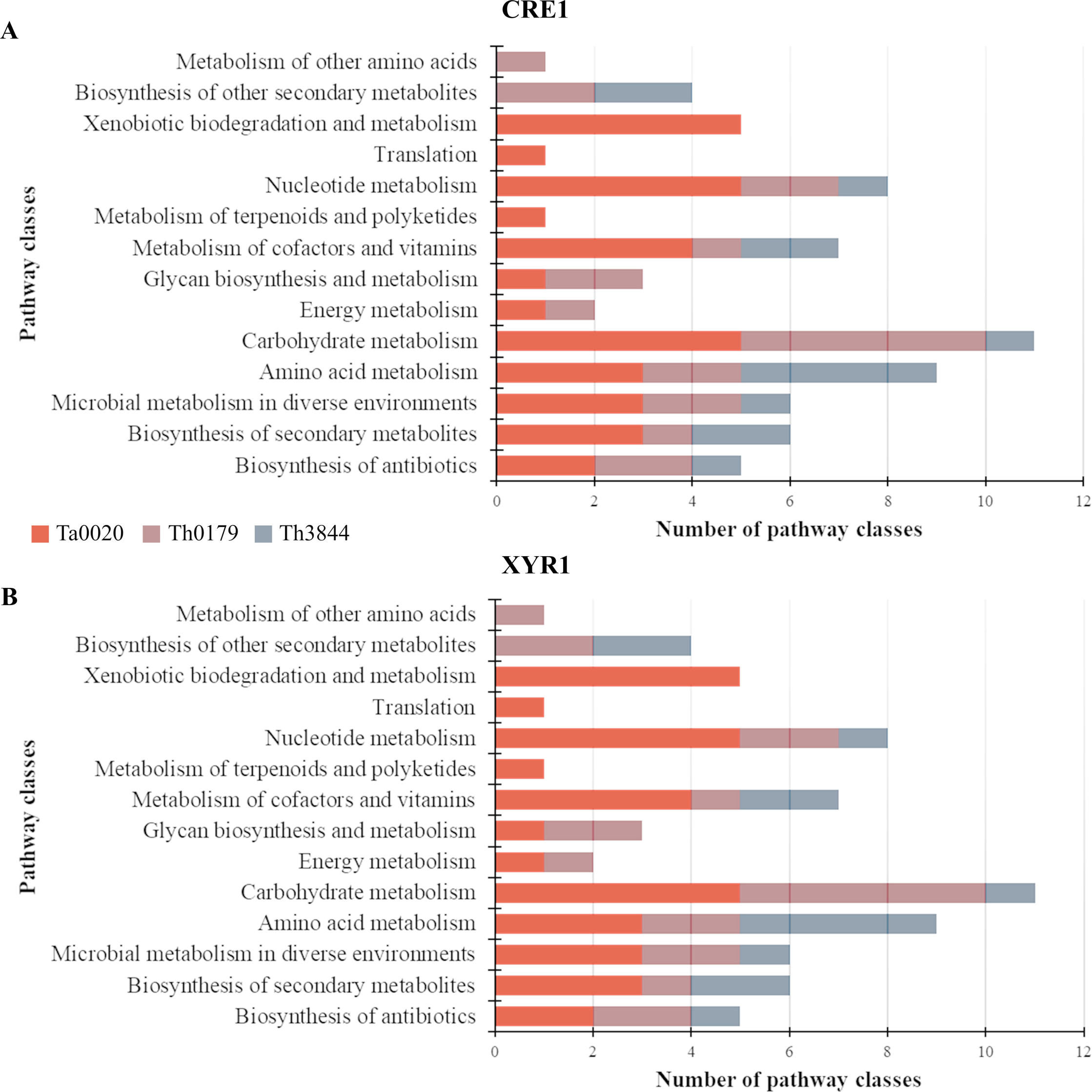
S2 Fig. Treemap Plotted Based on GO Categories of the Selected Groups for Th3844. GO analysis of all transcripts in the (A) *xyr1* group and (B) *cre1* group under the biological process category. Th3844: *T. harzianum* IOC-3844.

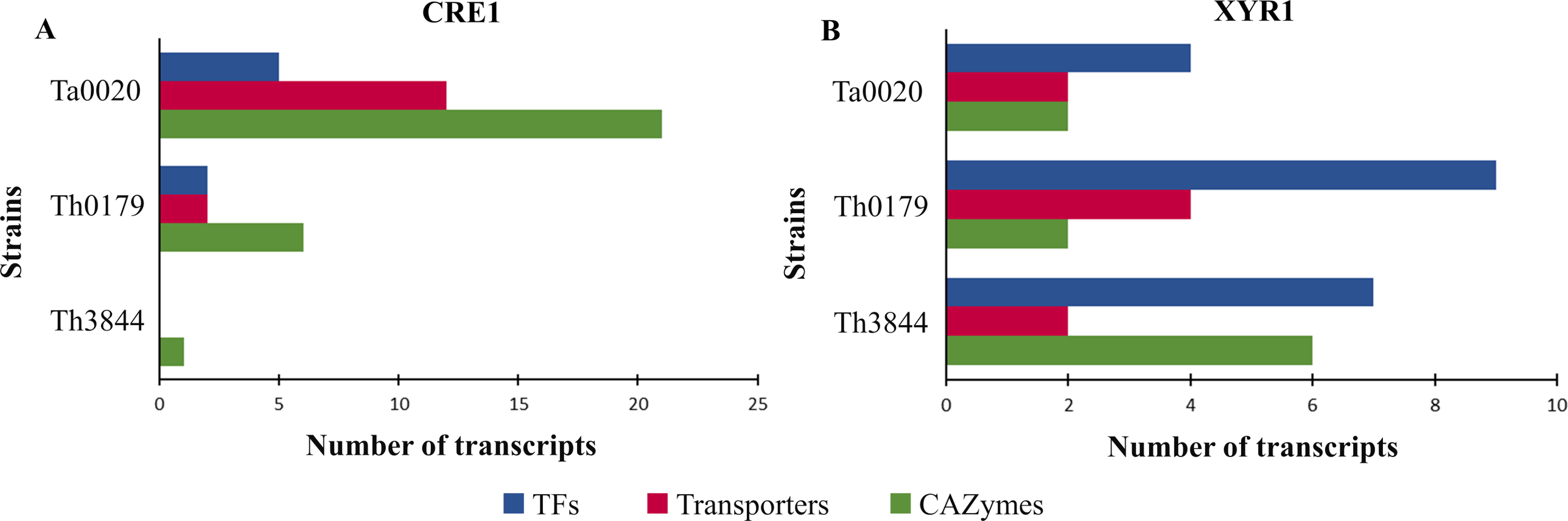
S3 Fig. Treemap Plotted Based on GO Categories of the Selected Groups for Th0179. GO analysis of all transcripts in the (A) *xyr1* group and (B) *cre1* group under the biological process category. Th0179: *T. harzianum* CBMAI-0179.

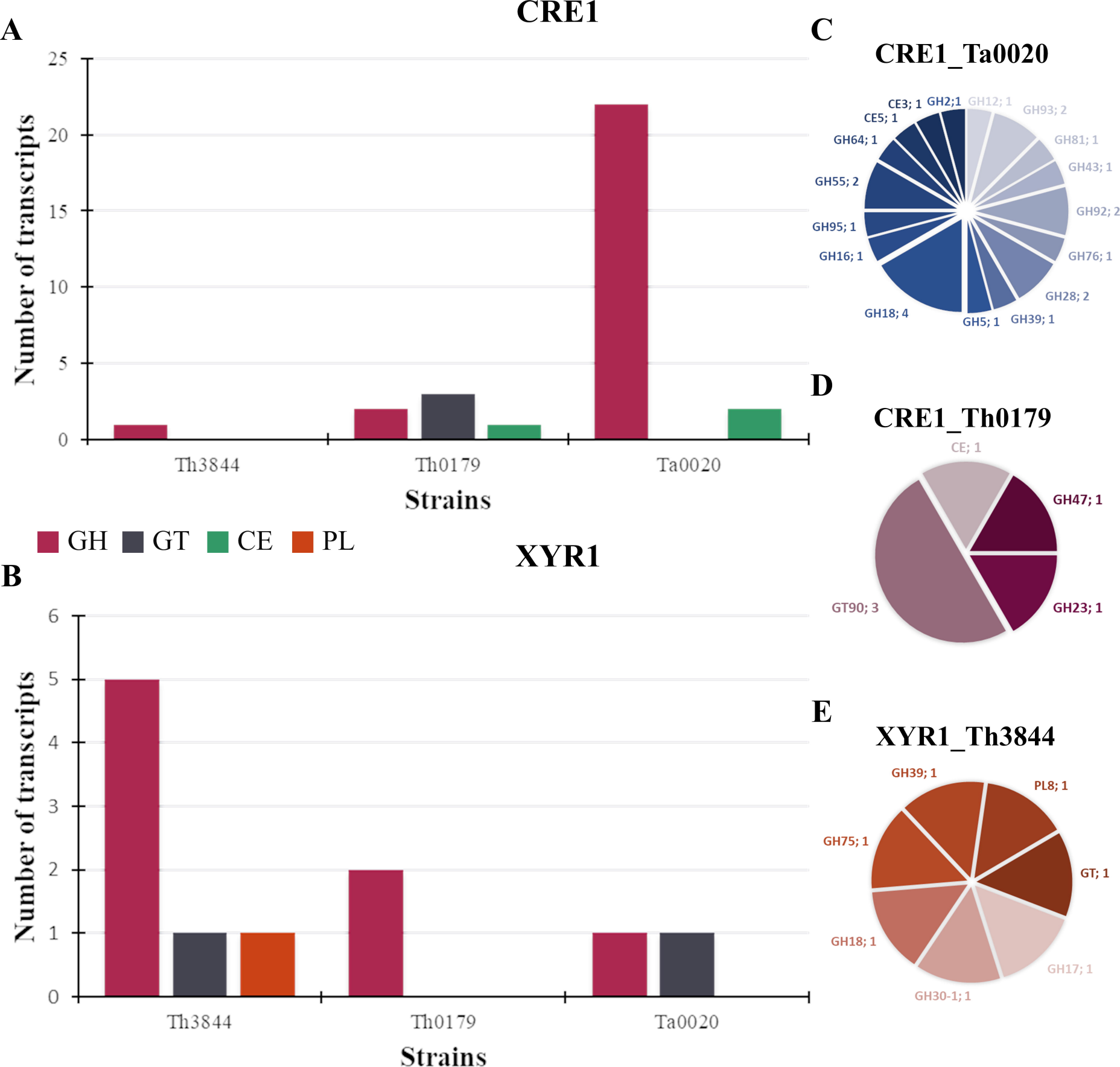
S4 Fig. Treemap Plotted Based on GO Categories of the Selected Groups for Ta0020. GO analysis of all transcripts in the (A) *xyr1* group and (B) *cre1* group under the biological process category. Ta0020: *T. atroviride* CBMAI-0020.

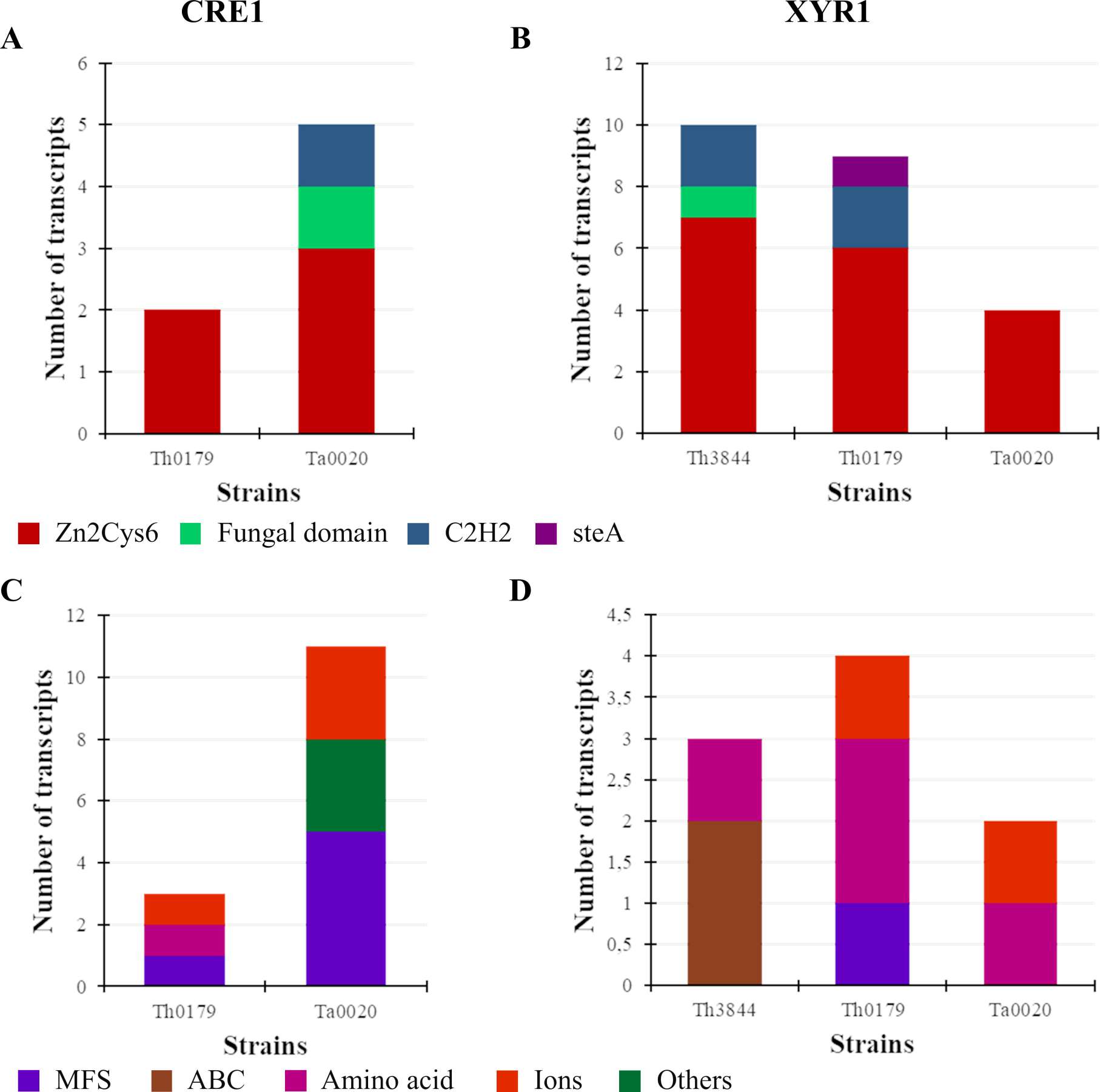

CRE1
XYR1
A
B
SORDARIACEAE
OPHIOCORDYCIPITACEAE
GLOMERELLACEAE
CLAVICIPITACEAE
NECTRIACEAE
MICROASCACEAE
HYPOCREACEAE
STACHYBOTRYACEAE
Incertae sedis
91-100
81-90
71-80


Th3844
Th0179
Ta0020
A
B
C
CRE1
D
E
F
XYR1


CRE1
A
Metabolism of other amino acids
Biosynthesis of other secondary metabolites
Xenobiotic biodegradation and metabolism
Translation
Nucleotide metabolism
Metabolism of terpenoids and polyketides
Metabolism of cofactors and vitamins
Glycan biosynthesis and metabolism
Energy metabolism
Carbohydrate metabolism
Amino acid metabolism
Microbial metabolism in diverse environments
Biosynthesis of secondary metabolites
Biosynthesis of antibiotics
Pathway classes
0
2
4
6
8
10
12
Number of pathway classes
Ta0020
Th0179
Th3844
XYR1
B
Metabolism of other amino acids
Biosynthesis of other secondary metabolites
Xenobiotic biodegradation and metabolism
Translation
Nucleotide metabolism
Metabolism of terpenoids and polyketides
Metabolism of cofactors and vitamins
Glycan biosynthesis and metabolism
Energy metabolism
Carbohydrate metabolism
Amino acid metabolism
Microbial metabolism in diverse environments
Biosynthesis of secondary metabolites
Biosynthesis of antibiotics
Pathway classes
0
2
4
6
8
10
12
Number of pathway classes


A
CRE1
Strains
Ta0020
Th0179
Th3844
0
5
10
15
20
25
Number of transcripts
B
XYR1
Strains
Ta0020
Th0179
Th3844
0
2
4
6
8
10
Number of transcripts
TFs
Transporters
CAZymes


A
CRE1
Number of transcripts
25
20
15
10
5
0
Th3844
Th0179
Ta0020
Strains
GH
GT
CE
PL
B
XYR1
Number of transcripts
6
5
4
3
2
1
0
Th3844
Th0179
Ta0020
Strains
C
CRE1_Ta0020
CE3; 1
GH2;1
GH12; 1
GH93; 2
CE5; 1
GH64; 1
GH81; 1
GH43; 1
GH55; 2
GH92; 2
GH95; 1
GH76; 1
GH16; 1
GH28; 2
GH18; 4
GH5; 1
GH39; 1
D
CRE1_Th0179
CE; 1
GH47; 1
GT90; 3
GH23; 1
E
XYR1_Th3844
GH39; 1
PL8; 1
GH75; 1
GT; 1
GH18; 1
GH17; 1
GH30-1; 1


CRE1
XYR1
A
B
C
D
Number of transcripts
Strains
Th0179
Ta0020
Th3844
Zn2Cys6
Fungal domain
C2H2
steA
MFS
ABC
Amino acid
Ions
Others