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

Rafaela Rossi Rosolen[a,b,1], Alexandre Hild Aono[a,b,2], Déborah Aires Almeida[a,b,3], Jaire Alves Ferreira Filho[a,b,4], Maria Augusta Crivelente Horta[a,5] and Anete Pereira de Souza[a,c*]

[a]Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, Avenida Cândido Rondon 400, Campinas, SP, 13083-875, Brazil

[b]Graduate Program in Genetics and Molecular Biology, Institute of Biology, UNICAMP, Campinas, SP, 13083-970, Brazil

[c]Department of Plant Biology, Institute of Biology, UNICAMP, Cidade Universitária Zeferino Vaz, Rua Monteiro Lobato, 255, Campinas, SP, 13083-862, Brazil

E-mail addresses

[1]rafaelarosolen@gmail.com

[2]alexandre.aono@gmail.com

[3]deborahairesa@gmail.com

[4]jaire_hp@yahoo.com.br

[5]mariaaugusta.horta@gmail.com

***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

Fax: +55-19-3521-1089

E-mail: anete@unicamp.br

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 explore the genetic mechanisms of both regulators by inferring a gene coexpression network for *T. harzianum* (IOC-3844 and CBMAI-0179) and for *T. atroviride* (CBMAI-0020). The phylogenetic analyses indicated that XYR1 and CRE1 are extensively distributed among ascomycete fungi and suggested how *T. atroviride* is differentiated from *T. harzianum*. Transcripts encoding carbohydrate-active enzymes (CAZymes), TFs, sugar and ion transporters, 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. The results differed according to the strains, suggesting that different molecular mechanisms related to XYR1 and CRE1 are used by *Trichoderma* spp. during cellulose degradation. These results can be exploited to improve the understanding of the regulation involved in hydrolytic enzyme expression, thereby expanding the potential of *T. harzianum* use in several industrial applications.

Keywords: *Trichoderma harzianum*, *Trichoderma atroviride*, cellulose, XYR1, CRE1, weighted correlation network analysis

Abbreviations

ABC: ATP-binding Cassette; **CAZymes:** Carbohydrate-active Enzymes; **CCR:** Carbon Catabolite Repression; **CEs:** Carbohydrate Esterases; **CRE1:** Carbon Catabolite Repressor 1; **GHs:** Glycoside Hydrolases; **GT:** Glycosyltransferase; **MCODE**: Molecular Complex Detection; **MFS:** Major Facilitator Superfamily; **PL:** Polysaccharide Lyase; **Ta0020:** *Trichoderma atroviride* CBMAI-0020; **TFs:** Transcription Factors; **Th0179:** *Trichoderma harzianum* CBMAI-0179; **Th3844:** *Trichoderma harzianum* IOC-3844; **WGCNA:** Weighted Correlation Network Analysis; **XYR1:** Xylanase Regulator 1.

1 Introduction

Many species of the filamentous ascomycete genus *Trichoderma* have been used in a wide range of biotechnological applications (Schuster and Schmoll, 2010). While fungal cellulases from *Trichoderma reesei* are the most applied enzymes in the biotechnology industry (Bischof et al., 2016), the use of *Trichoderma harzianum* and *Trichoderma atroviride* was explored by examining their biocontrol capacity against plant pathogenic fungi (Medeiros et al., 2017; Saravanakumar et al., 2017). 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 (Delabona et al., 2020; Zhang et al., 2020).

Lignocellulosic biomass is a complex recalcitrant structure that requires a consortium of enzymes for its complete depolymerization. Hydrolytic enzymes and other accessory proteins that can degrade the biomass compounds are collectively named carbohydrate-active enzymes (CAZymes) (Lombard et al., 2014). In *T. reesei*, the expression of CAZyme-encoding genes is regulated by the action of several transcription factors (TFs) (Portnoy et al., 2011b), and xylanase regulator 1 (XYR1) is the most important activator (Stricker et al., 2006). Moreover, the expression of these genes is subject to carbon catabolite repression (CCR), which is regulated by carbon catabolite repressor 1 (CRE1) (Strauss et al., 1995).

Although XYR1 orthologs are present in almost all filamentous ascomycete fungi, the molecular mechanisms triggered by this regulator depend on the species (Klaubauf et al., 2014). XYR1 was first described in *T. reesei* as the main activator of enzymes involved in d-xylose metabolism (Stricker et al., 2006). 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 (Reithner et al., 2014). CRE1 is the only conserved TF throughout the fungal kingdom, suggesting a conserved mechanism for CCR in fungi (Adnan et al., 2017), and in several cases, it has a negative effect on the production of lignocellulose-degrading enzymes (Alazi and Ram, 2018).

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. 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 (Pirim, 2018). To date, only a few studies have employed gene coexpression networks in studies on *Trichoderma* spp. (Almeida et al., 2020; Borin et al., 2018; Horta et al., 2018).

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. Such insights will be crucial to optimize the use of *T. harzianum* as an enzyme producer and expand the potential of this fungi in several industrial applications, including in the biofuel industry.

2 Materials and Methods

2.1 Fungal Strains, Culture Conditions and Transcription Profiling

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.

2.2 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 CBMAI and included in the phylogenetic analysis. 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 NCBI database. 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 included in the phylogenetic analyses, and the methodology is described in Additional file 1.

2.3 Gene Coexpression Network Construction

The gene coexpression networks were modeled for Th3844, Th0179, and Ta0020 using transcriptome data (PRJNA336221) (Horta et al., 2018), and Trimmomatic v0.36 (Bolger et al., 2014) was applied for sequencing quality filters (Almeida et al., 2020). The networks were constructed using transcripts per million (TPM) values with the R (R Core Team, 2018) package. Weighted correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) filters were applied and Pearson's correlation values 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 genes and edges correspond to the strength of their connection. To obtain a dissimilarity matrix, we built a topological overlap matrix as implemented in the package.

2.4 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 (Langfelder et al., 2007) 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), 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 (Sanner, 1999) together with the networkx library (Hagberg et al., 2008) to identify subgroups using the Louvain method (Blondel et al., 2008) and hubs based on the distribution degree. 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, we tested the Molecular Complex Detection (MCODE) (v1.4.2) plugin (Bader and Hogue, 2003) using Cytoscape software v3.7.2 with default parameters for subgroup identification.

2.5 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. All identified Gene Ontology (GO) (Ashburner et al., 2000) categories were used to create a treemap to visualize possible correlated categories in the dataset caused by

cellulose degradation using the REVIGO tool (Supek et al., 2011). The metabolic pathways related to the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) Orthology (KO) obtained and found 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 the different groups, we used the Python 3 programming language (Sanner, 1999) together with the BioPython library (Cock et al., 2009).

3 Results

3.1 Molecular Phylogeny of Trichoderma *spp.*

The strains studied were phylogenetically classified according to the ITS region (Additional file 1: Fig. S1). 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.

3.2 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: Table S1). The families observed in the CRE1 phylogenetic tree (Fig. 1a) were not the same as those identified for XYR1 (Fig. 1b).

3.3 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: Table S2). 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 (Additional file 4: Table S3).

3.4 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: Table S4). Each group containing the *xyr1* and *cre1* transcripts was partitioned, and subgroups were formed for all strains. The numbers of subgroups identified through MCODE algorithm are presented in Additional file 6: Table S5. The transcripts encoding CRE1 and XYR1 were only found for Th3844 (*cre1* and *xyr1*) and Th0179 (*cre1*). Then, we tested the Louvain algorithm (Fig. 2 and Additional file 7: Table S6).

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: Table S7). 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 subgroups formed for Th0179 and Ta0020 in the *cre1* group. 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 a few transcripts.

3.5 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: Table S8. The size of the correlation coefficient threshold variate to identify highly connected transcripts within a group (0.5 represents moderate correlation and 0.9 represents a relevant high correlation) (Mukaka, 2012).

3.6 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: Fig. S2, S3 and S4, 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 (Additional file 1: Fig. S5). For the 14 pathway classes observed for each TF, 13

were shared between the two regulators. However, the number of identified enzymes for a given pathway class was different for each strain.

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

Different quantities of differentially expressed transcripts were found for Ta0020 (78) and Th0179 (6) (described in Additional file 9: Table S9 and Additional file 1: Table S10). 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 (Antoniêto et al., 2014; dos Santos Castro et al., 2016). 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.

3.8 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. 3). Transcripts encoding TFs, transporters and CAZymes in the *cre1* and *xyr1* subgroups are described in Additional file 7: Table S6.

3.8.1 CAZymes

Glycoside hydrolases (GHs) were identified in all evaluated strains with different numbers (Additional file 1: Fig. S6). For the *cre1* group, while Ta0020 presented the highest number of classified transcripts from the GH family, 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). Carbohydrate esterases (CEs) were associated with the *cre1* transcript in Ta0020 and Th0179, while polysaccharide lyase (PL) was

coexpressed with the *xyr1* transcript in only Th3844. Glycosyltransferase (GT) was coexpressed with the *cre1* transcript in only Th0179 and with the *xyr1* transcript in Th3844 and Ta0020 (GT22).

3.8.2 Transcription Factors

Zn_2Cys_6 was the most abundant coexpressed transcript for the strains that showed transcripts encoding TFs in the *cre1* and *xyr1* groups (Additional file 1: Fig. S7). 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.

3.8.3 Transporters

Th3844 showed transcripts encoding ATP-binding cassette (ABC) transporters and amino acid transporters with a similar expression pattern as *xyr1* (Additional file 1: Fig. S8). In addition, all strains presented transcripts encoding amino acid transporters coexpressed with *xyr1*. For the *cre1* group, only Th0179 showed transcripts encoding this type of transport protein. In the *cre1* group, major facilitator superfamily (MFS) permeases were the most abundant proteins among the transporters for Ta0020. For the *xyr1* group, only Th0179 showed transcripts encoding this type of protein transport. Transcripts encoding ion transporters were present in the *cre1* and *xyr1* groups for Ta0020 and Th0179.

4 Discussion

The transcriptional regulation involved in the expression of CAZyme-encoding genes in *T. harzianum* species is still poorly explored when compared to that of the major producer of cellulases, *T. reesei*. In a transcriptomic study, were reported different types of enzymatic performance among *Trichoderma* species, in which Ta0020 had a lower cellulase activity profile during growth on cellulose than Th3844 and Th0179 (Almeida et al., 2020). Using these

transcriptome data, we chose to investigate and compare the genetic mechanisms involved in XYR1- and CRE1-mediated gene regulation through modeling the gene coexpression networks.

The diversification observed for CRE1 and XYR1 inside Ascomycota indicates a high functional diversity for these proteins that may reflect changes in their functions (Te Velthuis and Bagowski, 2008). For both regulators, a differentiation process was observed for Ta0020 compared to Th3844 and Th0179, which corroborates the molecular phylogeny of the species based on the ITS region. Furthermore, 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.

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.

Compared to Ta0020 and Th0179, these TFs had distinct metabolic processes 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 directly 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 (Mukherjee et al., 2008), was identified.

We also investigated the expression profile of the transcripts within the *cre1* and *xyr1* groups, including *xyr1* and *cre1*, under cellulose and glucose growth. A previous study observed that in the presence of glucose, the expression level of the transcript *xyr1* in *T. reesei* was minimal (dos Santos Castro et al., 2016), which was not observed herein for Ta0020. 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 (Benocci et al., 2017). Conversely, for Th3844, the *cre1* expression level was higher in the presence of glucose than in the presence of cellulose.

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 (Portnoy et al., 2011a). Transcripts encoding TFs and transporters were strongly repressed by CRE1 in *T. reesei* (Antoniêto et al., 2014), and only for Ta0020, we found several transcripts encoding these coexpressed proteins with the *cre1*.

For *T. reesei*, the expression of CAZymes, TFs, and transporters was affected by XYR1 in the context of plant biomass degradation (dos Santos Castro et al., 2016), whereas for *T. atroviride*, this TF was identified as a critical factor during the plant-fungus recognition processes and responsible for mycoparasitism activity (Reithner et al., 2014). 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 (Klaubauf et al., 2014).

Among CAZymes, GHs are the most abundant enzymes for the breakdown of glycosidic bonds between carbohydrate molecules (Cantarel et al., 2009). 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* (Gruber and Seidl-Seiboth, 2012). 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 to cellulose and hemicellulose degradation. The main CAZyme classes were not found in the *cre1* and *xyr1* groups for *T. harzianum*. It is possible that the transcripts encoding these hydrolytic enzymes were sorted in other groups formed within the coexpression networks, mainly because their expression patterns were different from those observed for the transcripts encoding XYR1 and CRE1.

Transporters have an important role in the utilization of lignocellulose by fungi (Nogueira et al., 2018). For *T. reesei*, MFS, amino acids and ABC transporters were modulated by XYR1 and CRE1 (Antoniêto et al., 2014; dos Santos Castro et al., 2016). 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* (dos Santos Castro et al., 2016). Conversely, the influence of this TF was less present in Ta0020, suggesting that both mycoparasitic *Trichoderma* spp. developed different mechanism about the import of molecules from the environment.

5 Conclusions

Our findings suggest that XYR1 and CRE1 triggered different signaling pathways in *Trichoderma* spp. during cellulose degradation and glucose metabolization, which could result in

the observed contrasting enzymatic performances during lignocellulose deconstruction. We expect that our results will contribute to a better understanding of the genetic mechanisms involved in hydrolytic enzyme regulation for *Trichoderma* species. This knowledge will be important for expanding the use of *T. harzianum* as an enzyme producer not only in the development of better enzyme cocktails but also in several other industrial applications.

Acknowledgements

We are grateful for 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.

Funding

Financial support for this work was provided by the São Paulo Research Foundation (FAPESP) (Process number 2015/09202-0 and 2018/19660-4), the Coordination of Improvement of Higher Education Personnel (CAPES, Computational Biology Program) (Process number 88882.160095/2013-01) and the Brazilian National Council for Technological and Scientific Development (CNPq) (Process number 312777/2018-3).

Role of the Funding Source

Not applicable.

Declarations of Interest

None.

Data Availability Statement

All data generated or analyzed during this study are included in this published article (and its supplementary information files). Raw RNA-Seq datasets are deposited at the NCBI Sequence Read Archive and can be accessed under the BioProject number PRJNA336221.

References

Adnan, M., Zheng, W., Islam, W., Arif, M., Abubakar, Y.S., Wang, Z., Lu, G., 2017. Carbon catabolite repression in *Filamentous* fungi. Int. J. Mol. Sci. 19, 48. https://doi.org/10.3390/ijms19010048.

Alazi, E., Ram, A.F.J., 2018. Modulating transcriptional regulation of plant biomass degrading enzyme networks for rational design of industrial fungal strains. Front. Bioeng. Biotechnol. 6, 133. https://doi.org/10.3389/fbioe.2018.00133.

Almeida, D.A., Horta, M.A.C., Ferreira Filho, J.A., Murad, N.F., de Souza, A.P., 2020. The synergistic actions of hydrolytic genes in coexpression networks reveal the potential of *Trichoderma harzianum* for cellulose degradation. bioRxiv 2020.01.14.906529 https://doi.org/10.1101/2020.01.14.906529.

Antoniêto, A.C.C., Castro, L.D.S., Silva-Rocha, R., Persinoti, G.F., Silva, R.N., 2014. Defining the genome-wide role of CRE1 during carbon catabolite repression in *Trichoderma reesei* using RNA-Seq analysis. Fungal Genet. Biol. 73, 93–103. https://doi.org/10.1016/j.fgb.2014.10.009.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29. https://doi.org/10.1038/75556.

Bader, G.D., Hogue, C.W.V., 2003. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinform. 4, 2. https://doi.org/10.1186/1471-2105-4-2.

Benocci, T., Aguilar-Pontes, M.V., Zhou, M., Seiboth, B., de Vries, R.P., 2017. Regulators of plant biomass degradation in *Ascomycetous* fungi. Biotechnol. Biofuels 10, 152. https://doi.org/10.1186/s13068-017-0841-x.

Bischof, R.H., Ramoni, J., Seiboth, B., 2016. Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. Microb. Cell Factories 15, 106. https://doi.org/10.1186/s12934-016-0507-6.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008, P10008. https://doi.org/10.1088/1742-5468/2008/10/p10008.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120. https://doi.org/10.1093/bioinformatics/btu170.

Borin, G.P., Carazzolle, M.F., dos Santos, R.A.C., Riaño-Pachón, D.M., Oliveira, J.V.D.C., 2018. Gene co-expression network reveals potential new genes related to sugarcane bagasse degradation in *Trichoderma reesei* RUT-30. Front. Bioeng. Biotechnol. 6, 151. https://doi.org/10.3389/fbioe.2018.00151.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. 37, D233–D238. https://doi.org/10.1093/nar/gkn663.

Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J.L., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25, 1422-1423. https://doi.org/10.1093/bioinformatics/btp163.

Delabona, P.D.S., Codima, C.A., Ramoni, J., Zubieta, M.P., de Araújo, B.M., Farinas, C.S., Pradella, J.G.D.C., Seiboth, B., 2020. The impact of putative methyltransferase overexpression on the *Trichoderma harzianum* cellulolytic system for biomass conversion. Bioresour. Technol. 313, 123616. https://doi.org/10.1016/j.biortech.2020.123616.

dos Santos Castro, L., de Paula, R.G., Antoniêto, A.C.C., Persinoti, G.F., Silva-Rocha, R., Silva, R.N., 2016. Understanding the Role of the Master Regulator XYR1 in Trichoderma reesei by Global Transcriptional Analysis. Frontiers in Microbiology 7, https://doi.org/10.3389/fmicb.2016.00175.

Gruber, S., Seidl-Seiboth, V., 2012. Self versus non-self: fungal cell wall degradation in Trichoderma. Microbiology 158, 26-34. https://doi.org/https://doi.org/10.1099/mic.0.052613-0.

Hagberg, A., Swart, P., Chult, D.S., 2008. Exploring network structure, dynamics, and function using networkx, in: Varoquaux, G., Vaught, T., Millman, J. (Eds.), Proceedings of the 7th Python in Science Conference (SciPy 2008). LANL, Los Alamos, US, pp. 11–16.

Horta, M.A.C., Filho, J.A.F., Murad, N.F., Santos, E.D.O., dos Santos, C.A., Mendes, J.S., Brandão, M.M., Azzoni, S.F., de Souza, A.P., 2018. Network of proteins, enzymes and genes linked to biomass degradation shared by *Trichoderma* species. Sci. Rep. 8, 1341. https://doi.org/10.1038/s41598-018-19671-w.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30. https://doi.org/10.1093/nar/28.1.27.

Klaubauf, S., Narang, H.M., Post, H., Zhou, M., Brunner, K., Mach-Aigner, A.R., Mach, R.L., Heck, A.J.R., Altelaar, A.F.M., de Vries, R.P., 2014. Similar is not the same: differences in the function of the (hemi-)cellulolytic regulator XlnR (Xlr1/Xyr1) in *Filamentous* fungi. Fungal Genet. Biol. 72, 73–81. https://doi.org/10.1016/j.fgb.2014.07.007.

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 9, 559. https://doi.org/10.1186/1471-2105-9-559.

Langfelder, P., Zhang, B., Horvath, S., 2007. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24, 719-720. https://doi.org/10.1093/bioinformatics/btm563.

Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495. https://doi.org/10.1093/nar/gkt1178.

Medeiros, H.A.D., Filho, J.V.D.A., Freitas, L.G.D., Castillo, P., Rubio, M.B., Hermosa, R., Monte, E., 2017. Tomato progeny inherit resistance to the nematode *Meloidogyne javanica* linked to plant growth induced by the biocontrol fungus *Trichoderma atroviride*. Sci. Rep. 7, 40216. https://doi.org/10.1038/srep40216.

Mukaka, M.M., 2012. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. 24, 69–71. https://doi.org/10.1038/75556.

Mukherjee, P.K., Nautiyal, C.S., Mukhopadhyay, A.N., 2008. Molecular Mechanisms of Biocontrol by Trichoderma spp, in: Nautiyal, C.S., Dion, P. (Eds.), Molecular Mechanisms of Plant and Microbe Coexistence. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 243-262.

Nogueira, K.M.V., de Paula, R.G., Antoniêto, A.C.C., dos Reis, T.F., Carraro, C.B., Silva, A.C., Almeida, F., Rechia, C.G.V., Goldman, G.H., Silva, R.N., 2018. Characterization of a novel sugar transporter involved in sugarcane bagasse degradation in *Trichoderma reesei*. Biotechnol. Biofuels 11, 84. https://doi.org/10.1186/s13068-018-1084-1.

Pirim, H., 2018. Construction of Gene Networks Using Expression Profiles, in: Purohit, H.J., Kalia, V.C., More, R.P. (Eds.), Soft Computing for Biological Systems. Springer Singapore, Singapore, pp. 67-89.

Portnoy, T., Margeot, A., Linke, R., Atanasova, L., Fekete, E., Sándor, E., Hartl, L., Karaffa, L., Druzhinina, I.S., Seiboth, B., Le Crom, S., Kubicek, C.P., 2011a. The CRE1 carbon catabolite repressor of the fungus *Trichoderma reesei*: a master regulator of carbon assimilation. BMC Genom. 12, 269. https://doi.org/10.1186/1471-2164-12-269.

Portnoy, T., Margeot, A., Seidl-Seiboth, V., Le Crom, S., Chaabane, F.B., Linke, R., Seiboth, B., Kubicek, C.P., 2011b. Differential regulation of the cellulase transcription factors XYR1, ACE2, and ACE1 in *Trichoderma reesei* strains producing high and low levels of cellulase. Eukaryot. Cell 10, 262–271. https://doi.org/10.1128/EC.00208-10.

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Reithner, B., Mach-Aigner, A.R., Herrera-Estrella, A., Mach, R.L., 2014. Trichoderma atroviride transcriptional regulator Xyr1 supports the induction of systemic resistance in plants. Appl. Environ. Microbiol. 80, 5274–5281. https://doi.org/10.1128/AEM.00930-14.

Sanner, M.F., 1999. Python: a programming language for software integration and development. J. Mol. Graph. Model. 17, 57–61. https://doi.org/10.1186/1471-2105-4-2.

Saravanakumar, K., Li, Y., Yu, C., Wang, Q.-Q., Wang, M., Sun, J., Gao, J.-X., Chen, J., 2017. Effect of *Trichoderma harzianum* on maize rhizosphere microbiome and biocontrol of fusarium stalk rot. Sci. Rep. 7, 1771. https://doi.org/10.1038/s41598-017-01680-w.

Schuster, A., Schmoll, M., 2010. Biology and biotechnology of Trichoderma. Applied microbiology and biotechnology 87, 787-799. https://doi.org/10.1007/s00253-010-2632-1.

Strauss, J., Mach, R.L., Zeilinger, S., Hartler, G., Stöffler, G., Wolschek, M., Kubicek, C.P., 1995. Crel, the carbon catabolite repressor protein from *Trichoderma reesei*. FEBS Lett. 376, 103–107. https://doi.org/10.1016/0014-5793(95)01255-5.

Stricker, A.R., Grosstessner-Hain, K., Würleitner, E., Mach, R.L., 2006. Xyr1 (xylanase regulator 1) regulates both the hydrolytic enzyme system and D-xylose metabolism in *Hypocrea jecorina*. Eukaryot. Cell 5, 2128–2137. https://doi.org/10.1128/EC.00211-06.

Supek, F., Bošnjak, M., Škunca, N., Šmuc, T., 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PLOS ONE 6, e21800. https://doi.org/10.1371/journal.pone.0021800.

Te Velthuis, A.J.W., Bagowski, C.P., 2008. Linking fold, function and phylogeny: a comparative genomics view on protein (domain) evolution. Curr. Genom. 9, 88–96. https://doi.org/10.2174/138920208784139537.

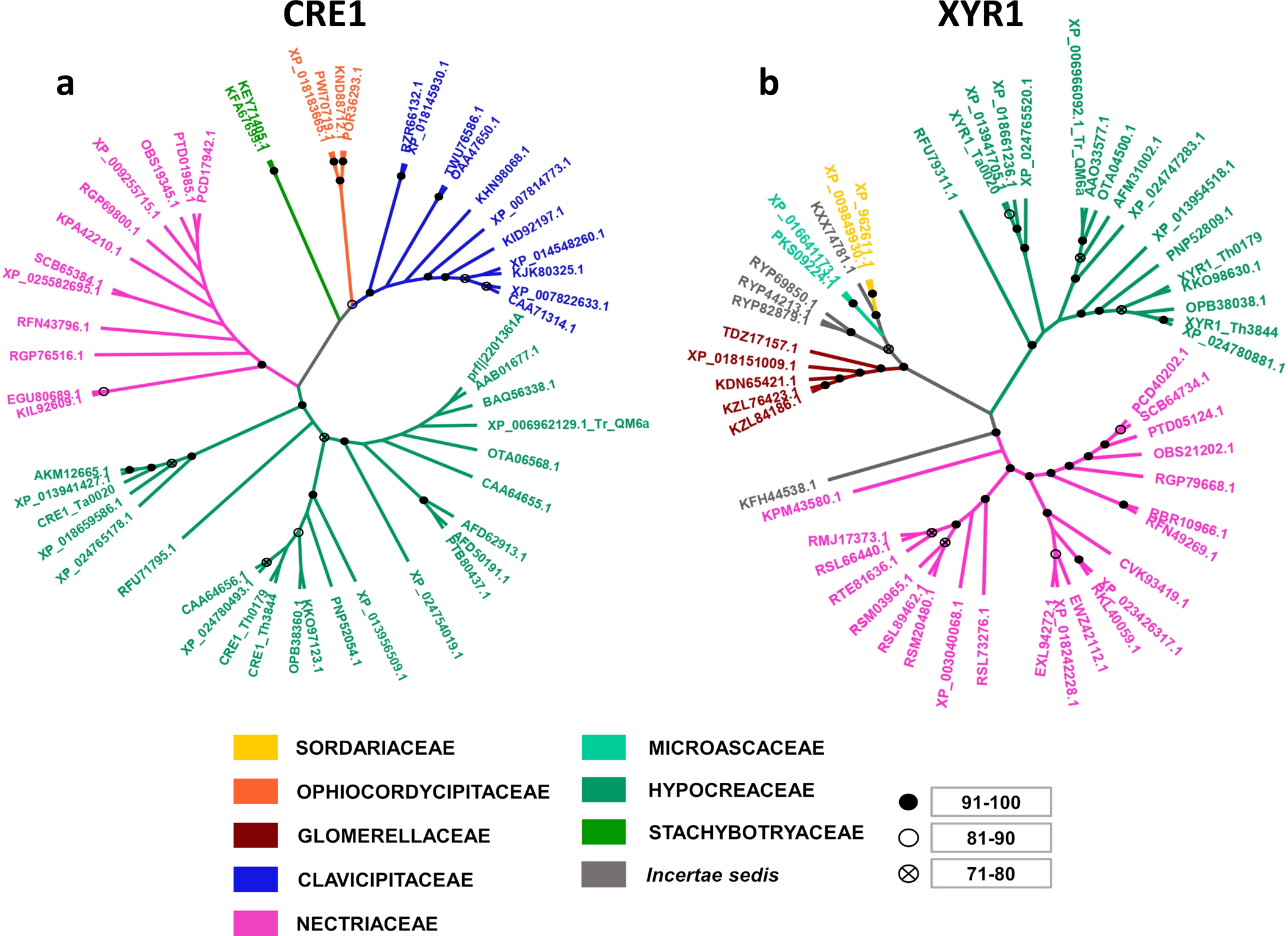
Zhang, Y., Yang, J., Luo, L., Wang, E., Wang, R., Liu, L., Liu, J., Yuan, H., 2020. Low-cost cellulase-hemicellulase mixture secreted by *Trichoderma harzianum* EM0925 with complete saccharification efficacy of lignocellulose. Int. J. Mol. Sci. 21, 371. https://doi.org/10.3390/ijms21020371.

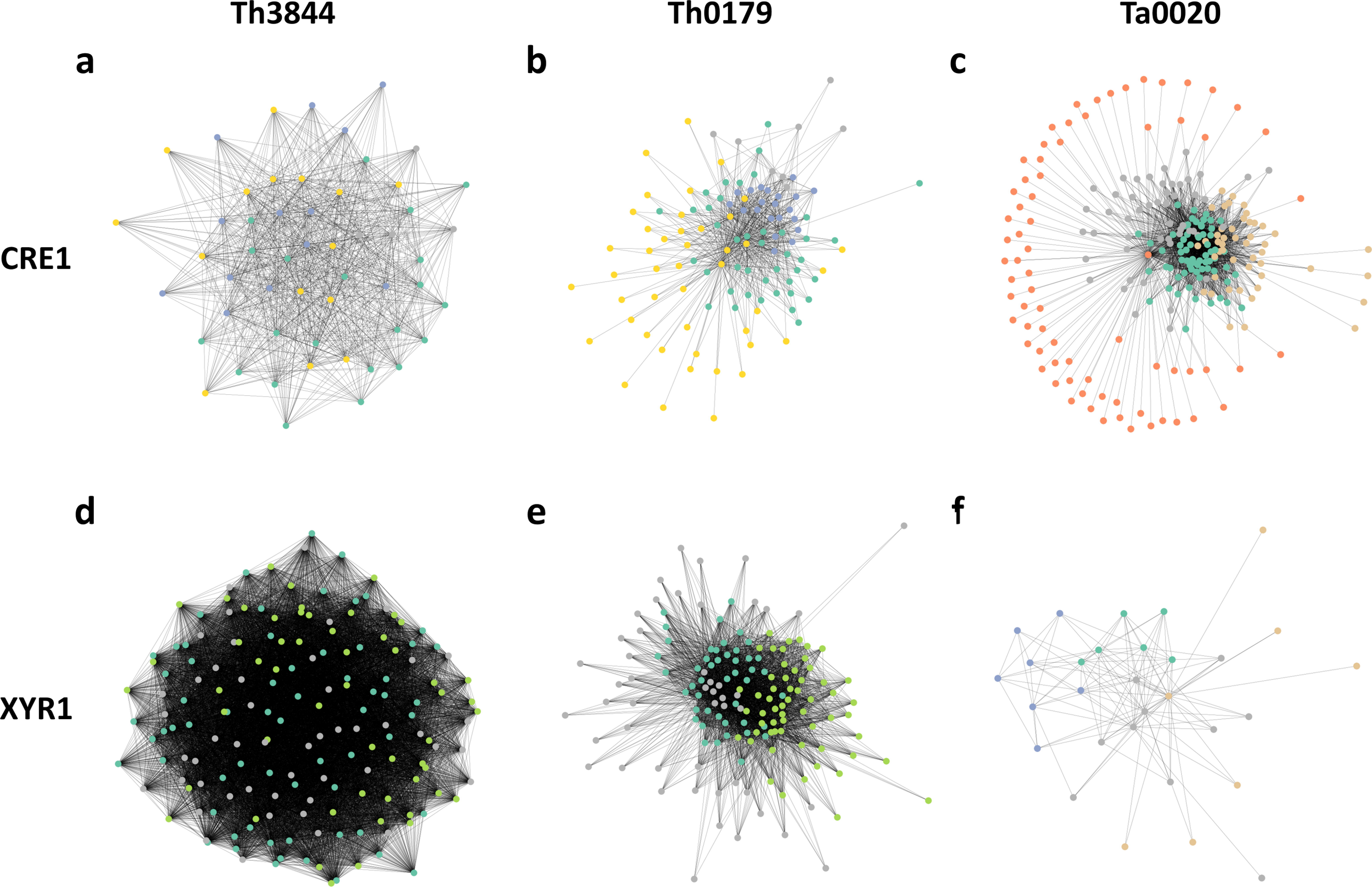
Figure Captions

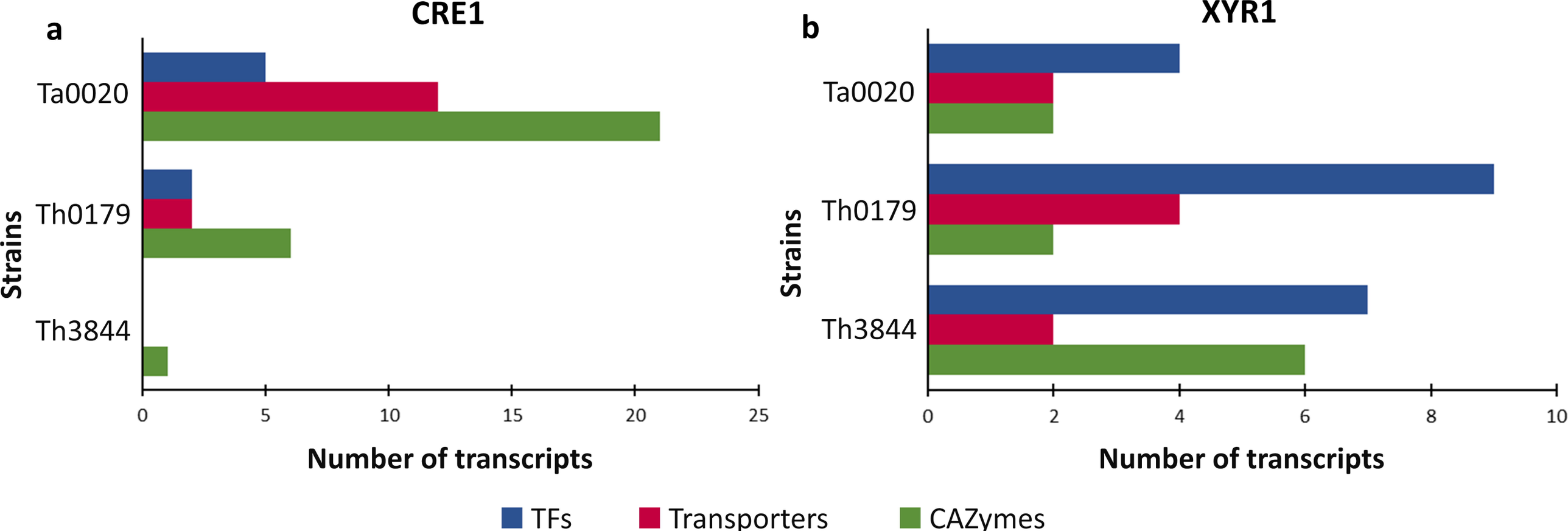
Fig. 1. Molecular phylogeny of CRE1 and XYR1 in Ascomycota. The complete protein sequences related to CRE1 (**a**) and XYR1 (**b**) were used to compare the phylogenetic relationships of the studied *Trichoderma* species/strains inside the ascomycete fungi. The phylogenetic trees were drawn using MEGA7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

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 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.

Fig. 3. TFs, transporters and CAZymes comparison 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. The number of transcripts encoding these proteins associated with *xyr1* were more distributed among all strains than those observed in the *cre1* group. TFs: Transcription factors; CAZymes: Carbohydrate-active enzymes. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844.


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


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