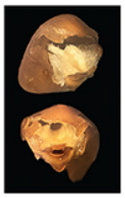
MYCOLOGIA



Mycologia

ISSN: 0027-5514 (Print) 1557-2536 (Online) Journal homepage: www.tandfonline.com/journals/umyc20

Targeting lignocellulolytic gene clusters in novel *Trichoderma atroviride* and *Trichoderma harzianum* strains through bacterial artificial chromosome–guided analysis

Paulo Henrique Campiteli, Maria Augusta Horta, Rafaela Rossi Rosolen, Juliano Sales Mendes, Carla Cristina da Silva, Danilo Sforça & Anete Pereira de Souza

To cite this article: Paulo Henrique Campiteli, Maria Augusta Horta, Rafaela Rossi Rosolen, Juliano Sales Mendes, Carla Cristina da Silva, Danilo Sforça & Anete Pereira de Souza (27 Jun 2025): Targeting lignocellulolytic gene clusters in novel *Trichoderma atroviride* and *Trichoderma harzianum* strains through bacterial artificial chromosome–guided analysis, Mycologia, DOI: 10.1080/00275514.2025.2496600

To link to this article: https://doi.org/10.1080/00275514.2025.2496600

MYCOLOGIA
https://doi.org/10.1080/00275514.2025.2496600

Taylor & Francis
Taylor & Francis Group

Targeting lignocellulolytic gene clusters in novel *Trichoderma atroviride* and *Trichoderma harzianum* strains through bacterial artificial chromosome–guided analysis

Paulo Henrique Campiteli[a,b], Maria Augusta Horta[a], Rafaela Rossi Rosolen[a,b], Juliano Sales Mendes[a], Carla Cristina da Silva[c], Danilo Sforça[a], and Anete Pereira de Souza[a,b,d]

[a]Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, Campinas, 13083-875, Brazil; [b]Graduate Program in Genetics and Molecular Biology, Institute of Biology, UNICAMP, Campinas, 13083-970, Brazil; [c]Department of Agronomy, Federal University of Viçosa, UFV, Av. P. H. Rolfs, Campus Universitário, 36570-900, Brazil; [d]Department of Plant Biology, Institute of Biology, UNICAMP, Cidade Universitária Zeferino Vaz, Rua Monteiro Lobato, Campinas, 13083-862, Brazil

ABSTRACT

Lignocellulosic biomass is a complex carbon source with recalcitrant properties whose degradation via industrial enzymatic hydrolysis is challenging, directly affecting the cost of reliable energy production. In nature, filamentous fungi, including *Trichoderma* species, degrade lignocellulose via an arsenal of hydrolytic and oxidative enzymes that act synergistically to process it into soluble sugar monomers. This work explored the genomic content of *Trichoderma atroviride* and *Trichoderma harzianum* strains with hydrolytic abilities by identifying regions possessing degradative enzyme–encoding genes, namely, hydrolytic clusters. We employed bacterial artificial chromosome (BAC) methodology to target specific genomic regions and explore their genetic organization, proximal gene context, and gene expression under degradative conditions. With this tool, it was possible to inspect the linear structure and expression profile of target hydrolytic-rich genomic regions. The present work offers a perspective on the organization of genome regions related to carbohydrate metabolism. This study revealed novel genes and genome regions that are positively regulated during cellulose degradation, contributing to elucidating differences in gene organization that potentially impact hydrolysis among *Trichoderma* species.

ARTICLE HISTORY
Received 26 September 2024
Accepted 13 April 2025

KEYWORDS
Bacterial artificial chromosomes (BACs); gene clusters; genomic organization; hydrolytic enzymes; *Trichoderma*

INTRODUCTION

Trichoderma spp. (ascomycetes) are opportunistic-generalist fungi that efficiently degrade plant biomass (Druzhinina et al. 2018; Kubicek et al. 2019). They are distributed worldwide and colonize soil, living plants, decomposing wood, and other fungi (Druzhinina et al. 2018; Kubicek et al. 2019; Srivastava and Shahid 2014; Zeilinger et al. 2016). These species, especially *Trichoderma reesei* QM6a (and its derivative clones), have previously been of interest in biotechnology due to their efficient lignocellulolytic activities (Li et al. 2017; Li and Wang 2021). They are widely used for the industrial production of cellulases and hemicellulases for enzymatic cocktails, which are applied in saccharification (Bischof et al. 2016) and are used to optimize enzymatic hydrolysis processes in biorefining, especially during the process of lignocellulose biodegradation. Biodegradation promoted by *Trichoderma* spp. is driven by a complex network of reactions, which first requires substrate perception, triggering specific metabolic reactions to synthesize and transport hydrolytic enzymes, classified as carbon-active enzymes (CAZymes), and other genes related to lignocellulolytic activity, such as transcription factors (TFs), transporters (TRs), as well as other complex metabolites (Andlar et al. 2018). CAZymes are enzymes related to cleavage, synthesis, or interactions with carbon polymers and are organized and classified into families and subfamilies according to the reactions in which they participate, as shown in the Carbon Active Enzymes database (http://www.cazy.org/) (Andlar et al. 2018; Drula et al. 2022).

In addition to their potential as a source of new and optimized hydrolytic enzymes, emerging interest in *Trichoderma* is linked to their observed biological activities, such as their biocontrol and biofertilization capabilities. Some species are capable of controlling rhizospheric

CONTACT Anete Pereira de Souza anete@unicamp.br
Supplemental data for this article can be accessed online at https://doi.org/10.1080/00275514.2025.2496600

Published online 27 Jun 2025

microorganisms while displaying beneficial interactions with plants and are considered prominent biocontrol agents (BCAs) that pose a low risk to humans (Kubicek et al. 2019; Sood et al. 2020). *Trichoderma atroviride* and *Trichoderma harzianum* are mycoparasites that exhibit biocontrol activity against various phytopathogens, including *Rhizoctonia* sp., *Botrytis* sp., *Fusarium* sp., and *Penicillium* sp. (Nascimento et al. 2022; Sood et al. 2020; Stracquadanio et al. 2020). Additionally, plant-*Trichoderma* cocultures exhibit biofertilization activities, promoting vegetative traits and immunological defense systems and attenuating abiotic stress (Guzmán-Guzmán et al. 2023; Kubicek et al. 2011, 2019; Sood et al. 2020; Tyśkiewicz et al. 2022).

The observed capabilities *Trichoderma* species indicate that members of this genus possess a complex system of substrate perception and are able to trigger specific reactions for metabolic adaptation by synthesizing and transporting proteins and complex metabolites for substrate degradation, defense, competition, and interaction with other organisms (Andlar et al. 2018). To explore the diversity in the genus and the differences in genome organization and CAZyme composition, we investigated *T. harzianum* CBMAI 179 (TH0179) and *T. atroviride* CBMAI 0020 (TA0020). Both species were previously investigated under biomass degradation conditions. The hydrolytic potential of these strains was evaluated by quantifying cellulase and hemicellulase enzymatic activities (Horta et al. 2018), and the transcriptomic profiles of both strains were explored by gene expression analysis under cellulose degradation conditions (Almeida et al. 2021). TH0179 exhibited increased enzymatic activity under cellulose degradation conditions, whereas TA0020 presented a more complex transcriptomic profile with decreased lignocellulolytic activity (Almeida et al. 2021; Horta et al. 2018; Rosolen et al. 2022). The observed heterogeneous responses to cellulose degradation indicate the high complexity of gene expression at the genus level, wherein closely related species can carry relevant species-specific traits that affect the biological response of the fungus.

The bacterial artificial chromosome (BAC) is a molecular tool used to recover and study target genomic regions (Peterson et al. 2000). It is a selection platform that distributes the entire genome in libraries, which allows the identification of specific genomic region targets by polymerase chain reaction (PCR). This technique can be applied for gap closure in fragmented assemblies, as an assembly method (as described for the first human genome sequence; Waterston et al. 2002)) or as a tool for studying targeted genomic regions in higher eukaryotes with complex genomes rich in polyploid and aneuploid repetitive elements (Mancini et al. 2018; Visendi et al. 2016). It is a simple, reliable, low-cost method for studying target genomic regions and is especially powerful for studying novel species without previous genomic data. Among fungi, the BAC approach has been successfully applied to investigate genomic regions with CAZyme cluster genes in *T. harzianum* IOC 3844 (Crucello et al. 2015; Filho et al. 2020). In *Streptomyces* spp., for example, the BAC approach was used for genome mining and revealed important gene clusters for bioactive natural products and the corresponding biosynthetic pathways (Xu et al. 2016). Another study used BAC libraries to characterize a novel gene conferring sensitivity to the necrotrophic fungal pathogen *Parastagonospora nodorum* (Zhang et al. 2021).

In this study, the BAC approach was applied to investigate hydrolytic activity–related genomic regions in the TH0179 and TA0020 strains. This methodology allowed us to describe genomic regions enriched with coding cellulases, hemicellulases, and ligninases, as well as TFs, transporters, and potential targets involved in cellulose degradation. In addition, we employed the data generated from a previous RNA-Seq study, HiSeq 2500 platform (Illumina, San Diego, CA, USA) (Almeida et al. 2021) to analyze the expression of CAZyme genes through a BAC-guided analysis, which allowed the identification of lignocellulolytic CAZyme clusters that are positively regulated for the degradation of cellulose.

The results offer insights into the genomic organization and gene clustering of hydrolytic enzymes in the studied strains and were integrated with the transcriptomic data. The results revealed differences in lignocellulolytic activities between the species. Additionally, species-specific genomic regions and gene targets were described for future characterization.

MATERIALS AND METHODS

Strains and reference sequences.—TH0179 and TA0020 were obtained from the Brazilian Collection of Environment and Industry Microorganisms (CBMAI), located in the Multidisciplinary Center for Chemical, Biological, and Agricultural Research (CPQBA), Campinas, São Paulo, Brazil. The identities of the *Trichoderma* isolates used in these studies were authenticated by a provider on the basis of phylogenetic studies of their internal transcribed spacer (ITS) region, translational elongation factor 1 (*tef1*) marker gene, and RNA polymerase II (*rpb2*) marker gene (Kimura et al., 1980; Raeder et al., 1985; Saitou et al., 1987; Tamura et al. 2007; Thompson et al. 1994). A phylogenetic tree based on the ITS region of this species was previously published (Rosolen et al. 2022).

As the genomes of the studied species were not available at the time of the experiments, *T. harzianum* T6776 (Baroncelli et al. 2015) was used as a reference genome for TH0179 and *T. atroviride* IMI206040 (Kubicek et al. 2011) was used as a reference genome for TA0020. The reference genomes provided the consensus sequences for the target gene primer design and for RNA-Seq read mapping. The transcriptomes of TH0179 and TA0020 cultivated with crystalline cellulose (inducing degradative conditions) or glucose (control conditions) as a unique carbon source were previously described (National Center for Biotechnology Information [NCBI] BioProject PRJNA336221) (Almeida et al. 2021).

Our experimental results (Almeida et al. 2021; Rosolen et al. 2022) have consistently shown the importance of studying novel strains and exploring their specific response to biological processes such as the lignocellulolytic activities of *Trichoderma*. The goal of this report is to review similarities and differences and identify novel relevant genetic players that can be used in industrial settings to optimize existing industrial applications. The integration of bioinformatic tools and modern multiomics characterization offers an opportunity for strain engineering to obtain more efficient phenotypes (Benocci et al. 2018; Salazar-Cerezo et al. 2023).

BAC library construction and screening.—For each species, 10 µL of spore culture was inoculated on potato dextrose agar (PDA) plates (100 µg/mL ampicillin, 34 µg/mL chloramphenicol) at 28 C for 4 days. Next, a 2 cm × 2 cm portion of the colony was inoculated on sterilized potato dextrose and incubated (28 C, 3 days, 200 rpm). The final liquid culture was filtered using sterile Miracloth (Calbiochem, San Diego, USA), and the mycelia were subjected to high-molecular-weight genomic DNA (HMW-gDNA) extraction using the Qiagen DNeasy Plant Mini Kit II (Hilden, Germany). The HMW-gDNA was immobilized on low melting-point agarose plugs for BAC library construction, and large genomic DNA inserts were cloned and inserted into *E. coli* using the pAGI BAC vector (provided by the French Plant Genomic Resources Center [CNRGV], National Research Institute for Agriculture, Food and Environment [INRAE], France). Inserts were transformed via electroporation into *E. coli* DH10B competent cells.

The transformed culture was inoculated on PDA plates, and the successfully inserted fragments were identified visually as white colonies. Each individual colony represented a unique clone and was cultured and plated on 384-well plates. Each library was examined on a rapid selection platform that pooled the samples in rows or columns, which allowed fast screening of the library through real-time quantitative polymerase chain reaction (qPCR) of target genes using a previously described method (Crucello et al. 2015; Filho et al. 2020). The sequences of the primers used were designed on the basis of target CAZymes (Untergasser el al., 2012). The entire BAC library consisted of 5760 individual clones, each carrying a unique long genomic fragment (20 to 200 kb), constructed as described previously (Peterson et al. 2000), with minor modifications for filamentous fungal samples.

Screening of CAZyme genes and BAC-DNA extraction.—The primers were used for amplification on the selection platform, and BACs were selected by melting curve comparison. To avoid nonspecific and false-positive amplification, the final selection was performed by 3% agarose electrophoresis (80 V, 2 h, 1 kB Promega, Madison, WI, USA). The selected clones were preincubated in growth medium (15 mL, 200 rpm, 37 C, 6–12 h) and then incubated overnight in 80 mL of growth medium (200 rpm, 37 C, 6–12 h). After incubation, the culture was centrifuged and BAC-DNA was extracted using a Macherey-Nagel NucleoSpin Plant II mini kit (Düren, Germany) following standard procedures.

BAC-DNA assembly and evidence-based gene prediction and expression.—The selected BACs were sequenced by the Arizona Genomics Institute (PacBio RS II, Pacific Biosciences, Menlo Park, CA, USA). The sequencing data were transferred to a local server for polishing and assembly. The raw reads were polished using Quiver, and the BACs were de novo assembled using Canu 2.1.1 (Koren et al. 2017). The assembled genomic regions were aligned to the *T. harzianum* T6776 (Baroncelli et al. 2015) and *T. atroviride* IMI 206040 (Kubicek et al. 2011) reference genomes to identify *Trichoderma* target sequences from the respective bulk BAC sequences. BACs <40 kb in size were filtered out before analysis. Then, all-vs.-all alignment was performed for the TA0020 and TH0179 BACs individually to select nonoverlapping regions. For structural comparison, the BACs were aligned to the *T. atroviride* IMI 206040 (Kubicek et al. 2011) and *T. harzianum* CBS226.95 (Baroncelli et al. 2015) draft genomes and the *T. reesei* QM6a (Li et al. 2017) reference genome using MUMmer (Marçais et al. 2018). The MUMmer dot plot parameters employed to visualize the alignment

were as follows: all-vs.-all, minimum match 50 bp, and minimum alignment length 100 bp.

For evidence-based gene prediction, we first aligned RNA-Seq reads to the selected BACs using STAR (Dobin et al. 2013). BAM outputs were input into StringTie 2.2.1 (Pertea et al. 2015) for transcript assembly guided by the BAC sequences. The final set of transcripts was used as input for MAKER2 2.31.11 (Holt and Yandell 2011) for gene structure prediction. The predicted models were used for further functional inference and comparative analysis. Gene models were processed with InterProScan 5.55-88.0 (Jones et al. 2014; Quevillon et al. 2005) and eggNOG-mapper 2.0.1 (Huerta-Cepas et al. 2019). Parallel annotation of modules and clusters coding CAZymes was conducted with the dbCAN3 Web platform (Zheng et al. 2023). BlastKOALA (Kanehisa et al. 2016) was used to identify metabolism-related genes from http://www.kegg.jp (Kanehisa and Goto 2000) via a search against Hypocreceae. The antiSMASH Web server 6.0 (Blin et al. 2021) was used to define secondary metabolite biosynthesis–related gene clusters (SMBGCs). The functional data and gene structure predictions were integrated using Funannotate (Li and Wang 2021; Palmer and Stajich 2020), which was implemented on the Galaxy platform (Galaxy Community 2022).

The expression values of genes present in the genomic regions of TH0179 and TA0020 selected by BAC-guided assembly were extracted from previously published data (Almeida et al. 2021) using FastQC 0.11.5 (Trivedi et al. 2014) and Trimmomatic 0.39 (Bolger et al. 2014) with the following parameters: a minimum length of 36 bp and a SLIDINGWINDOW ratio of 4:10:10 were used to assess and filter the reads by quality. Filtered reads were aligned to BAC sequences with STAR 2.7.10b aligner (Dobin et al. 2013), and StringTie 2.2.1 (Pertea et al. 2015) was used for transcript assembly. In addition, the transcripts per million (TPM) value for each gene was calculated automatically on the basis of StringTie quantification. The EDGER package (Robinson et al. 2010) was used to compare gene expression under the studied conditions and identify genes with significant differential expression.

To identify orthologous genes, OrthoFinder 2.5.2 (Emms and Kelly 2019) was used to compare newly annotated gene models among 24 *Trichoderma* species, with *Saccharomyces cerevisiae* S288C (Goffeau et al. 1996), *Neurospora crassa* OR74A (Galagan et al. 2003), *Pleurotus ostreatus* PC15 (Riley et al. 2014), and *Phanerochaete chrysosporium* RP-78 (Ohm et al. 2014) as external groups, all of which were extracted from the mycocosm (Grigoriev et al. 2014).

RESULTS

Target BAC selection.—The final BAC library consisted of 5760 individual clones, each carrying a unique long genomic fragment (ranging from 20 to 200 kb), and was examined using a rapid selection platform with plates containing pools for rapid screening of the whole library. For genomic region screening, 42 and 54 BAC clones were selected for TH0179 and TA0020, respectively. Each positively extracted clone was subjected to quality control, and the integrity profile of the pooled samples was checked via a fingerprint test (HindIII BAC digestion) involving 1% agarose electrophoresis and 3% agarose electrophoresis. After BAC sequencing, all-vs.-all sequence alignment was performed for each strain. Only BACs larger than 40 kb were considered for further analysis. After filtering, a final set of 24 nonredundant genomic regions for TH0179 and 44 for TA0020, ranging in size from 41 to 138 kb and 53 to 164 kb, respectively, was obtained. BACs were submitted to the NCBI database via BankIt under BioProject PRJNA1028979, Biosamples TH0179 SAMN37846728 and TA0020 SAMN37846658 (TABLE 1).

The size of the TH0179 genomic region was 1.9 Mb, approximately 5% that of the *T. harzianum* CBS226.95 genome. On the other hand, 44 unique regions were recovered from TA0020, reaching 4.7 Mb in size, which was 15% the size of the *T. atroviride* IMI 206040 genome. A comparative analysis of the generated sequences with reference genomes revealed that the alignment rate against *T. reesei* QM6a was 12.3% for TH0179 and 7.5% for TA0020. A total of 95.5% of the TH0179 sequence was aligned with the *T. harzianum* CBS226.95 reference, and 87.3% of the TA0020 sequence was aligned with

Table 1. Genomic and functional characterization overview.

Parameter	TH0179	TA0020
Sequenced BACs	24	44
Sequenced length (bp)	1955378	4721491
Mean BAC length (bp)	81474.08	107306.61
GC (%)	49.64%	49.94%
Predicted gene models	629	1280
Average gene length (bp)	1704.13	1953.11
Multiexon transcripts	615	1267
Complete CDS	589	1207
Interpro	492	1054
Gene ontology	404	889
eggNOG	542	1138
Pfam	437	948
Cazy	59	142
SignalP	71	131
Merops	20	60
Kegg	181	432
Secondary metabolism core enzymes	4	7

*Abbreviations: CDS: Coding DNA Sequence; Pfam: Protein Families database; Cazy: Carbon active enzymes data base; signalP: Signal peptide; MEROPS: peptidase database; Kegg: Kyoto Encycloped of Genes.

TaIMI 206040 as a reference. Notably, there as variation in gene position among species; FIG. 1 shows the distribution of the aligned sequences in the different reference genomes, with the presence of orthologs in different genomic regions, despite a tendency toward similar gene organization that increased with species similarity.

To establish orthologous relationships, we analyzed 30 fungal species plus the gene models generated in this study. In total, 310 808 genes were analyzed, 94.6% of which were assigned to 19 020 orthogroups, including 1392 species-specific orthogroups. Forty-eight orthogroups were shared among all of the strains. A total of 1343 (92.2%) TA0020 genes were assigned

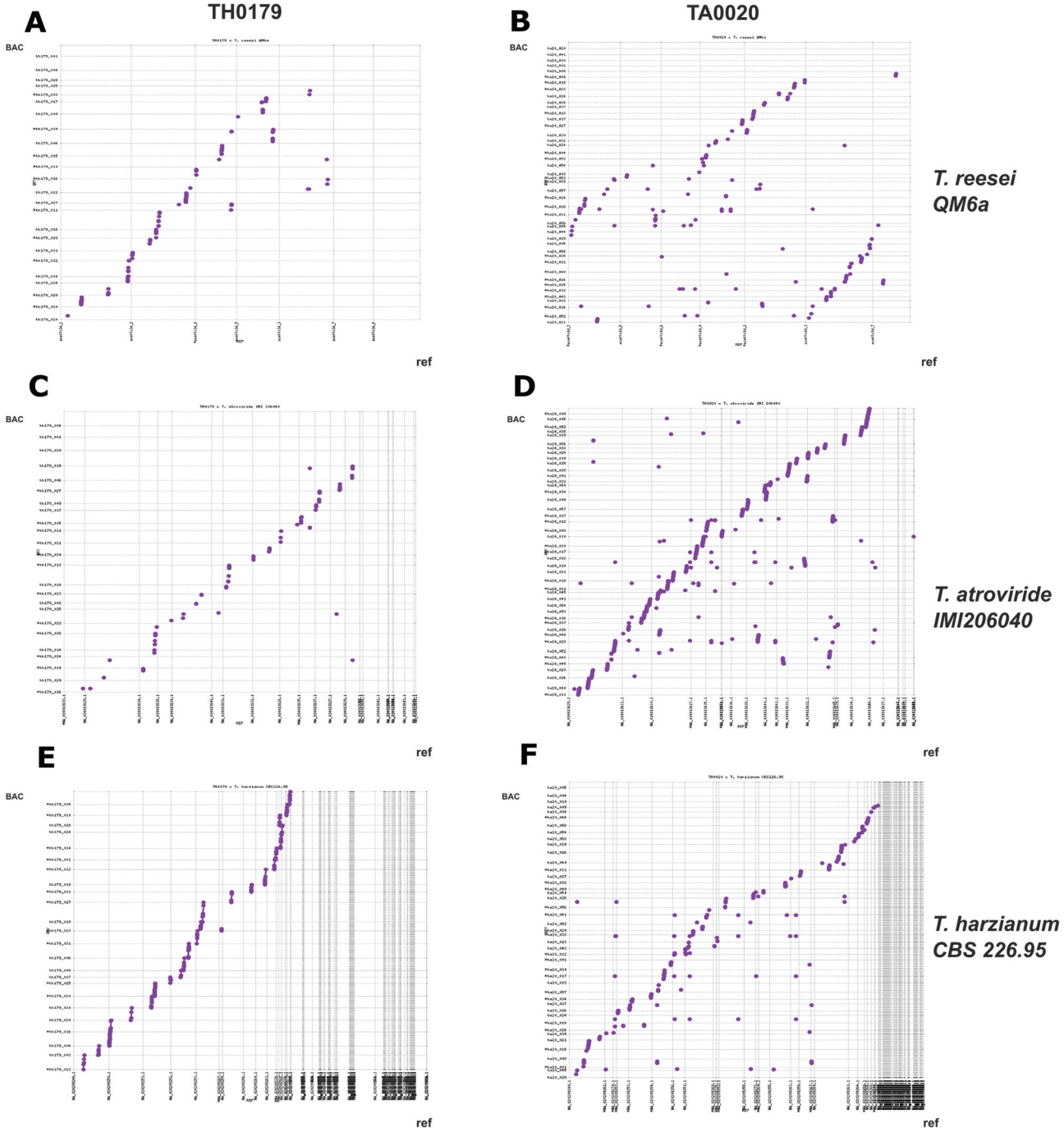


Figure 1. Alignments, visualized using a MUMmer dot plot with the parameters Minmatch 50 bp and match length 500 bp, against *T. reesei* QM6a, *T. harzianum* CBS22695, and *T. atroviride* IMI 206040. A. TH0179 × *T. reesei* QM6a. B. TA0020 × *T. reesei* QM6a. C. TH0179 × *T. atroviride* IMI206040. D. TA0020 × *T. atroviride* IMI206040. E. TH0179 × *T. harzianum* CBS226.95. F. TA0020 × *T. harzianum* CBS226.95.

to orthogroups, with 37 genes remaining unassigned, and they were distributed in almost all 24 genomic regions. Among the TH0179 genes, 596 (94.8%) were assigned to orthogroups and 33 remained unassigned and distributed in 17 genomic regions. In addition, 168 orthogroups were shared between TH0179 and TA0020, including genes encoding 15 CAZymes, 12 TRs, 10 TFs, and 5 SMBs.

Multiple tools were used to access functional data for the annotated models. Ninety-three percent of the TA0020 annotated genes and 91% of the TH0179 annotated genes had at least one functional annotation. CAZyme modules represented 9.3% of the TH0179 annotated gene models and 11% of the TA0020 annotated gene models.

CAZymes, transporters, and transcription factors and clusters.—In total, 59 CAZymes, 27 TFs, and 39 TRs were identified in TH0179. Seventy-one genes presented signal peptides, including 23 CAZymes, one TR, one TF, and two SMGs, which were linked to the NRPS core enzyme from OR633216. Forty-two percent of the annotated CAZymes had predicted substrates, four of which had multiple targets. With respect to CAZyme clustering, in TH0179, three genomic regions harbored no CAZymes and five regions presented only one each. The other 15 selected regions had two to six CAZymes, among which OR633232 was considered the BAC with the AA2, CE4, GH5, CE4, and AA11 CAZymes; interestingly, all of these regions were related to cellulose degradation to some extent, with the GH and CE genes considered lignocellulolytic, whereas the AA redox family was involved in complex substrate degradation. dbCAN3 can also be used to identify signature CAZyme clusters on the basis of genomic localization and homology to known clusters in the database. For TH0179, 14 regions presented annotated CAZyme clusters.

For TA0020, 143 CAZymes, 62 TFs, and 114 TRs were identified. A total of 131 genes presented signal peptides, including 43 CAZyme genes, two TF genes, two TR genes, and five SM genes. Forty percent of the CAZymes of TA0020 had predicted substrates, with 13 showing more than one target substrate. With respect to CAZyme clustering, four regions did not present any CAZymes, four regions presented one CAZyme each, and 36 genomic regions presented two or more CAZymes. Additionally, 31 regions contained dbCAN3-annotated CAZyme clusters.

Overall, the TA0020 gene models presented a greater number of annotated CAZymes, TFs, and TRs than did TH179, but for both strains it was possible to identify regions in which those hydrolytic genes were near the same genome position, in tandem with other genes predicted to be involved in metabolic hydrolysis control.

Expression analysis under cellulose degradation conditions.—Among all the TA0020 BACs, 62 genes were significantly differentially expressed (DEGs), 35 of which were upregulated when cellulose was used as a carbon source for culture, whereas 27 were downregulated.

The downregulated Differently expressed genes p value less than 0.05 and logFC greater than 1.5 or lesser than -1.5 included GH17 (TA0020_0933), AA9 (TA0020_1125), CBM18 (TA0020_1121), three TRs (TA0020_0935, TA0020_0744, TA0020_1208), six TFs (TA0020_0369, TA0020_0374, TA0020_0371, TA0020_0358, TA0020_0735, TA0020_1197), an additional biosynthetic SM gene from the T1PKS SMBGC cluster (TA0020_0056), and a direct neighbor gene from the terpene core-enzyme SMBGC (TA0020_1004). The cellulose-related DEGs included GH34 (TA0020_0446), GT1 (TA0020_1099), GT20 (TA0020_1261), and GH28 (TA0020_0170); three TFs (TA0020_1100, TA0020_0593, and TA0020_0023); two genes from the Ripp-like SMBGC; one gene with no annotation (TA0020_0494); and one gene encoding a protein with an alpha/beta hydrolase fold domain (TA0020_0489). Among the DEGs, the functional inference for two downregulated and five upregulated genes remained unclear, and these genes are therefore potential new targets for future characterization.

Among the 14 DEGs identified in all of the TH0179 BACs, three were downregulated and 11 were upregulated; these DEGs were distributed in nine genomic regions. The downregulated genes were TH0179_0577 (eggNOG hit with a hypothetical protein ortholog), TH0179_0308 TF (Zn(II)2-Cys6 fungal type), and TH0179_0309 TR Major Facilitator Superfamily (MFS). The 10 upregulated genes included three genes encoding CAZymes: TH0179_0102 (GH18), TH0179_0045 (AA7), and TH0179_0001 (GH95). Additionally, TH0179_0393 and additional biosynthetic genes from SMBGCs, TH0179_0359 (tetratricopeptide repeat, a protein-protein interaction driver), TH0179_0589 (WSC domain, a carbohydrate-binding domain), TH0179_0302 (PAN domain), TH0179_0217 (only an eggNOG hypothetical protein hit), TH0179_0053 TR (related to cation transport), and TH0179_0248 (CorA-like Mg^{2+} transporter protein), were identified. AA7 also belongs to an annotated CAZyme cluster in the OR633223 genomic region, presented in the next section (FIG. 3).

Major degradation-related regions and clusters.— For in-depth investigation of the genome regions captured by the BACs, the gene clusters that presented degradative genes were modeled, and the gene position, annotation, and/or expression information is shown in FIG. 3 for TA0020 and FIG. 4 for TH0179.

By crossing the DEGs and the CAZyme clusters, it was possible to identify major cellulolytic and glycolytic CAZyme clusters among the annotated BACs. For TA0020, eight relevant genomic regions were identified.

OR643395 carried three CAZymes. The first gene in the annotated CAZyme cluster was a downregulated gene, TA0020_0615 (a flavin reductase related to redox activities). OR643400 carried three CAZymes and multiple TFs and TRs. Within the cluster, two TFs, TA0020_0737 and TA0020_0745, which are both Zn(II)2-Cys6 fungal-type TFs, were upregulated. One gene, TA0020_0751 (alginate lyase), was downregulated. OR643406 also carried three CAZymes; the first gene from the CAZyme cluster, TA0020_0927 (pyridoxal phosphate–dependent enzyme), a direct neighbor of GH3, was upregulated. OR643407 carried two CAZymes, both in the annotated CAZyme cluster. TA0020_0445 (ferric reductase nicotinamide adenine dinucleotide [NAD]-binding domain) and TA0020_0473 (no functional hit) were upregulated. In contrast, TA0020_0447 (carboxylesterase family, bifunctional feruloyl and acetyl xylan esterase [BD-FAE], a novel bifunctional carbohydrate esterase [CE]), TA0020_0450 TF (Zn(II)2-Cys6 fungal type), and TA0020_0453 TR (ABC transporter and transmembrane region domain) were downregulated. OR643408 carried six CAZymes, all of which were in the annotated CAZyme cluster. Four genes were upregulated: TA0020_0234 (GH20), TA0020_0239 (taurine catabolism dioxygenase TauD, TfdA family), TA0020_0251 TR (peptidase merops family M28, Major Facilitator Superfamily (MFS)), and TA0020_0252 (taurine catabolism dioxygenase TauD, TfdA family). OR643409 carried two CAZymes, both in the CAZyme cluster. Two genes were upregulated: TA0020_0203 (amino-transferase class IV) and TA0020_0211 (amidohydrolase family). OR643425 carried a CAZyme (GH26) in the annotated clusters, plus two others at the edge of the region. Within the clusters, the gene TA0020_0678 (fungal pheromone mating factor STE2 G protein–coupled receptor [GPCR] and indoleamine 2,3-dioxygenase domains) was upregulated. Outside the cluster, the gene TA0020_0689 TR Major Facilitator Superfamily (MFS) was downregulated. OR643427 did not present any CAZyme. Additionally, seven genes were downregulated, one of which was TA0020_0380 (amino acid permease). Interestingly, the other six downregulated genes belonged to the annotated SMBGC, including the nonribosomal peptide synthetase (NRPS)-like core biosynthetic enzyme TA0020_0395 (adenosine monophosphate [AMP]-binding enzyme, phosphopantetheine attachment site, transferase), TA0020_0396 (pyoverdine/dityrosine biosynthesis protein, sporulation related), TA0020_0397 (enoyl-(acyl carrier protein) reductase), TA0020_0398 (male sterility protein), TA0020_0400 (2OG-Fe(II) oxygenase superfamily, nonheme dioxygenase in morphine synthesis N-terminal), and TA0020_0401 (DUF218 domain).

For TH0179, nine regions presented DEGs, among which three contained CAZyme clusters with upregulated genes. One region without CAZymes but carrying three DEGs was identified.

OR633211 carried five CAZymes, all in the annotated CAZyme cluster. In addition, one SMBGC was identified in the same region. The expression of the gene TH0179_0393 (zinc-binding dehydrogenase), which belongs to the SMBGCs, was upregulated. OR633219 carried no CAZyme. The expression of the TH0179_0302 gene (PAN domain, which mediates binding to high-molecular-weight (HMW) kininogen) was upregulated. TH0179_0308 TF (Zn(II)2-Cys6 fungal type) and TH0179_0309 TR Major Facilitator Superfamily (MFS) were downregulated. OR633223 carried a CAZyme cluster with five CAZymes, and, interestingly, an SMBGC overlapped in the region. TH0179_0045 (AA7), belonging to both functional clusters, and TH0179_0053 TR (haloacid dehalogenase–like hydrolase) were upregulated. OR633224 carried four CAZymes, three of which were in the annotated cluster. TH0179_0001 (GH95), which was not part of the cluster, was upregulated.

Overall, relevant genomic regions with CAZyme-annotated genes and clusters were observed. The expression analysis of TH0179 and TA0020 helped to identify major players in degradative conditions that were likely active under the conditions tested herein and were integrated with other genes in their vicinities.

DISCUSSION

BAC-guided transcriptome analysis was used to mine high-quality genomic regions containing CAZyme targets with biotechnological relevance from TA0020 and TH0179. This methodology promoted *hybrid* evidence-based and ab initio gene prediction, followed by functional inference and alignment-based expression analysis to

identify regions potentially linked to cellulose and glucose degradation. Overall, the main goal was to inspect the genomic and transcriptomic variations and review the genomic context of potential CAZymes and other relevant features related to the cellulolytic metabolism process. A range of clustered CAZymes, including cellulases, hemicellulases, and ligninases, were identified. Additionally, important genes, such as those encoding SMBGCs, oxidases, proteinases, TFs, and TRs, were closely related to CAZymes in the gene sequence clusters. Additionally, genes with high expression levels under degradative conditions were identified and are likely relevant to the degradative process. The exact relationships of the clustered genes, how their expression is orchestrated to respond to lignocellulose synergistically, and the species-specific particularities affecting the ligninolytic process were reviewed from a nonmodel species perspective for *Trichoderma* species.

Genomic region characterization.—The BAC approach is a methodology that can be used to successfully target specific regions and guarantee the quality and stability of specific genomic regions; thus, this method avoids assembly mistakes, especially in repetitive regions of the genome. This work explored specific regions of the genome; however, the total length obtained for TA0020 was greater than that for TH179, and the differences in the number of genes identified between the two strains can be explained by this limitation (FIG. 2). Therefore, the BAC sequences obtained and compared with the genomes highlight the diversity within the genus (FIG. 1). Previous comparative analysis revealed that 50–25% of the *Trichoderma* genes are species-specific genes and are not shared among *Trichoderma* species. Additionally, gene families such as the ankyrin, heterokaryon incompatibility domain protein (HET), TF, carbohydrate degradative, and SMBGC genes all presented high polymorphism rates (Kubicek et al. 2011, 2019; Mukherjee et al. 2013). The intraspecific diversity of the *Harzianum* clade is a challenge to modern molecular genetics because this species aggregates multiple highly complex infraspecific groups (Druzhinina et al. 2010; Kubicek et al. 2019; Schalamun and Schmoll 2022), which can partially explain the limited CAZyme gene transfer.

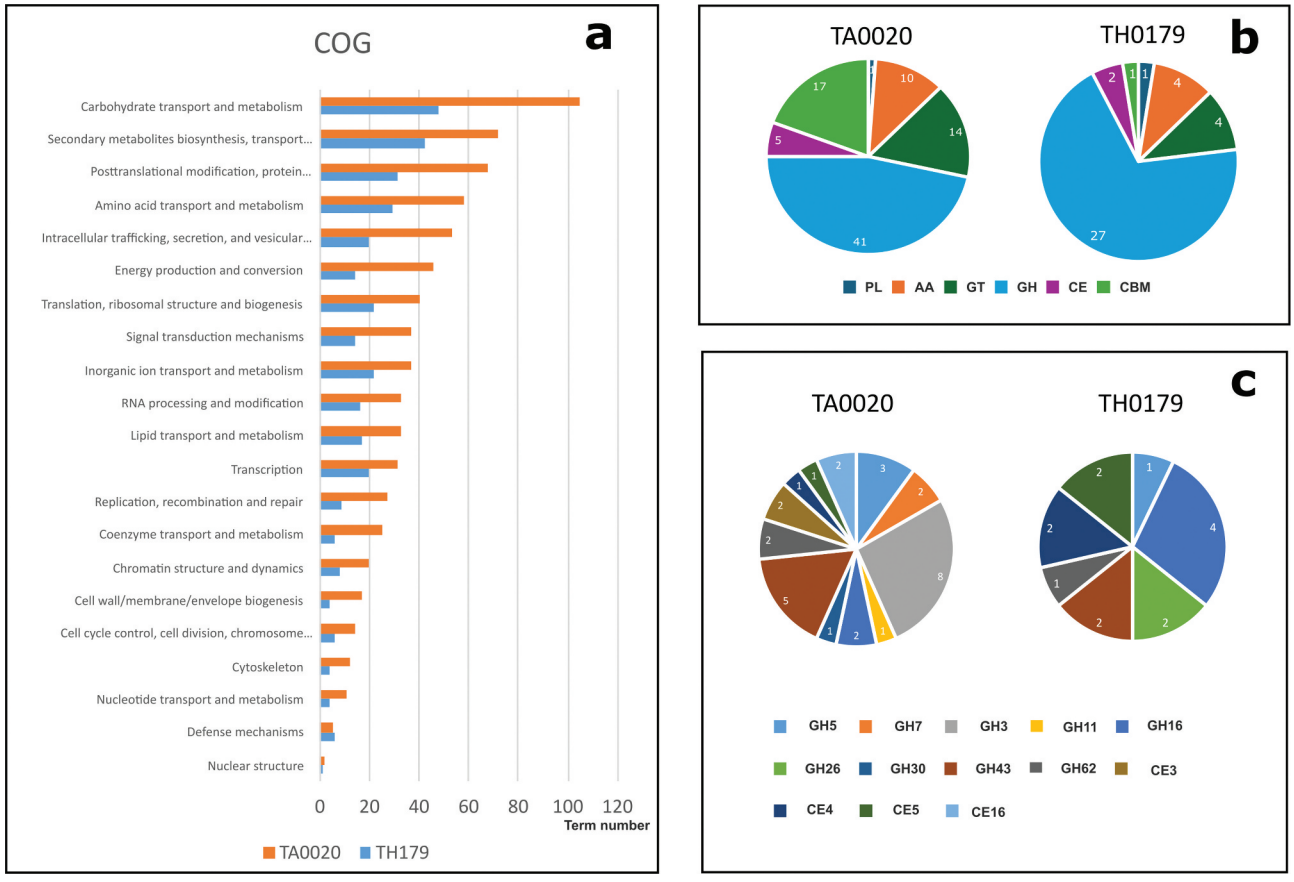


Figure 2. a. COG categories attributed to annotated BAC genes. b. CAZyme families annotated in each strain. c. Lignocellulolytic CAZymes (Andlar et al. 2018) annotated for each strain. The numbers of classified genes for each class are shown in the plot.

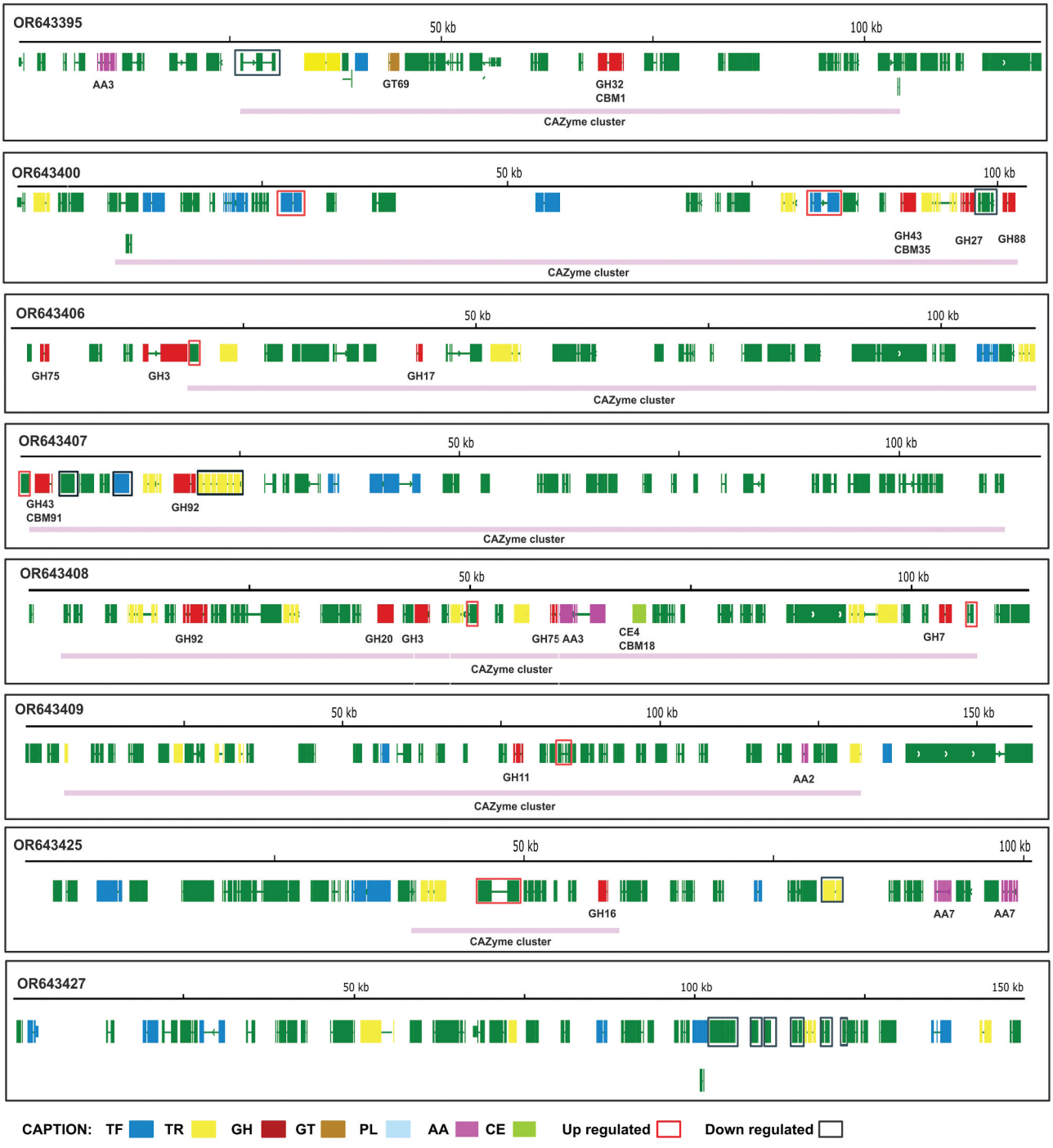


Figure 3. TA0020 Integrative Genome Viewer (IGV) visualization of CAZyme clusters or DEG-carrying regions identified via functional inference, genomic context, and differential expression analysis.

Comparative analysis.—The BAC alignment with the reference genome and orthology data helped in the assessment of the complexity of *Trichoderma* at the strain level. *T. reesei* QM6a is considered a model *Trichoderma* species that is phylogenetically distant from other biocontrol species (Kubicek et al. 2019). As expected, the BAC correspondence decreased with phylogenetic distance. On the basis of the alignment to its own species reference (FIG. 1), it is possible to conclude that TA0020 may be more distant from *T. atroviride* IMI206040 than TH0179 is from *T. harzianum* CBS226.95. Additionally, for all three comparisons, the TA0020 alignments presented greater complexity, with the same BAC aligning to multiple contigs/chromosomes in the references. In the orthology analysis, only 33 TH0179 genes were unassigned to orthogroups,

Figure 4. TH0179 Integrative Genome Viewer (IGV) visualization of CAZyme clusters or DEG-carrying regions identified via functional inference, genomic context, and differential expression analysis.

whereas 37 genes for TA0020 remained unassigned. According to a previous molecular phylogenetic analysis of *rpb2, T. atroviride* species belong to a basal branch that diverged early from other species, most likely maintaining the characteristics of a common ancestor (Kubicek et al. 2019).

Overall, the analysis of the BAC gene regions corroborated the findings of rapid evolution rates due to the opportunistic lifestyle of *Trichoderma*. The rapid evolution of the genus reshaped its genomes, allowing these species to colonize potential niches (Kubicek et al. 2019). Their notable intraspecific capacities are strongly influenced by lateral gene transfer and other complex evolutionary processes contributing to genomic variations, even those observed in closely related species. Interestingly, carbohydrate metabolism and biocontrol gene families seem to be especially influenced by evolutionary processes (Guzmán-Guzmán et al. 2023; Hewedy et al. 2020; Kubicek et al. 2011, 2019; Schalamun and Schmoll 2022; Sood et al. 2020; Tyśkiewicz et al. 2022). Previous transcriptome and coexpression network analyses have shown differential efficiency and a high degree of variability at the strain level (Almeida et al. 2021; Horta et al. 2018; Rosolen et al. 2022). Further exploration of *Trichoderma* capabilities could help identify specific traits that are useful for second-generation bioethanol production and other industrial processes that use biocatalysts derived from genes and recombinant proteins. (Lima et al. 2024).

CAZyme genes and clusters.—CAZymes play a central role in carbon substrate degradation, energy consumption, cell wall synthesis, and other accessory activities mediated by substrate availability and other environmental stimuli (Drula et al. 2022).

For the studied species, the retrieved CAZyme genes represent a small subset of all CAZymes (Filho et al. 2017; Kubicek et al. 2011; Rosolen et al. 2023). Whole-genome analysis has shown that *T. reesei* genomes contain fewer CAZymes than do those of biocontrol strains (Kubicek et al. 2011; Li et al. 2017; Li and Wang 2021). This is related to the fact that *T. reesei* is a saprophytic species that has lost biocontrol-related gene families, including important CAZyme genes (Kubicek et al. 2019), and the recent whole-genome comparative

analysis of *T. harzianum* IOC 3844, TH0179, and TA0020 and *T. reesei* 711 corroborated these differences (Rosolen et al. 2023). For the retrieved genomic regions in the present study, TA0020 presented a diverse set of cellulases, hemicellulases, and ligninases despite its lower lignocellulolytic potential and lower expression of CAZyme genes, which may be related to plant-*Trichoderma* interactions partially degrading the plant cell wall, allowing the exchange of nutrients in endophytic species.

Complex substrate degradation, such as that of glucose, cellulose, lignocellulose, and chitin, activates different and complex synergistic reactions with different sets of genes, including CAZymes, TFs, and TRs. The clustered distribution of genes contributing to the same phenotype is not new, but in eukaryotes genes are expressed individually, and the clustering of metabolic genes in the fungal genome is still unclear (Rokas et al. 2018; Slot 2017; Wisecaver and Rokas 2015). In terms of the clustered distribution of CAZymes, overall, more than half of the CAZyme-containing BACs presented multiple CAZymes in tandem with important regulators, transporters, and other genes with a variety of functions (FIG. 3 and 4). Careful inspection of such clusters must be conducted to improve the understanding of the coregulatory patterns and synergistic actions among clustered genes, clarifying the lignocellulolytic activities of *Trichoderma* spp. Species-specific differences in the CAZyme arsenal, linear organization, and cluster organization lead to differential efficiency. This knowledge can be used to engineer strains via classic mutagenesis or CRISPR (clustered regularly interspaced short palindromic repeats)-Cas-9 models, modulating specific phenotypes or more efficient strains (Filiatrault-Chastel et al. 2021; Kun et al. 2019). Thus, proper genetic modifications can be directly applied in bioprocesses, such as the use of pretreatments with optimized enzymatic cocktails or biorefinery approaches. Overall, genetic modifications can influence the process of biomass degradation by optimizing access to sugars and other biomass-related compounds, increasing the cost-effectiveness of lignocellulolysis-based fermentation and increasing the reliability of bioprocesses at the industrial level (Filiatrault-Chastel et al. 2021; Raulo et al. 2021; Sardi and Gasch 2017).

Expression measurement and identification of differently expressed genes.—A prevalidated RNA-Seq (Almeida et al. 2021) data set was used as a source to identify regions/genes potentially activated during lignocellulose degradation. These previous studies characterized the global transcriptome through transcriptomic profiling, revealing different strategies for carbon utilization (Almeida et al. 2021). Transcriptomic, genomic, and phenotypic variations are common even at the strain level, and such differences must be carefully studied to understand their role in the biodegradation process (Almeida et al. 2021; Hewedy et al. 2020; Pachauri et al. 2023; Rosolen et al. 2023). In TA0020 and TH0179, genes presented differences in expression, but few genes presented significant differences. These genes are likely regulated in response to the tested conditions; thus, these studies reveal the genomic context of these genes (FIG. 3 and 4).

The complex transcriptomic profile of TA0020, observed in the previous de novo transcriptome analysis, was captured in this reference-based transcriptome assembly and differential expression analysis. Their lower lignocellulolytic potential contrasts with their complex transcriptome profile, likely because *T. atroviride* may not be as specialized in lignocellulose utilization as the saprophytic workhorse *T. reesei* QM6a or the *T. harzianum* IOC 3844 and *T. harzianum* CBMAI-0179 strains, which show greater efficiency in crystalline cellulose degradation (Almeida et al. 2021; Horta et al. 2018; Li and Wang 2021). Transcriptomic and differential expression analyses offer a global understanding of complex activities and are reliable for identifying relevant target genes and regulatory and metabolic pathways. Such knowledge from a nonmodel and novel species perspective is important for synthetic biology and strain engineering efforts and heavily impacts biodegradation-related industrial processes that use *Trichoderma*-based technology (Benocci et al. 2018; Salazar-Cerezo et al. 2023; Xu et al. 2024).

CONCLUSIONS

The BAC methodology allowed us to identify CAZyme-containing genomic regions and study their linear distribution, genomic context, and relationship with degradative conditions. It was possible to identify target lignocellulolytic CAZyme clusters and address the differential lignocellulolytic potential in terms of gene expression previously observed among species. Functional inference combined with transcript expression quantification and comparative analysis provided highly valuable information regarding the genome arrangement of CAZyme genes, making it possible to identify strain-specific characteristics. We recovered multiple CAZyme-containing regions rich in TF- and TR-flanking genes, revealing the control mechanisms for CAZyme expression. Expression analysis helped to

identify regions that were positively regulated under biomass degradation conditions and likely involved in the lignocellulolytic activities of both strains.

These results contribute to the genomic perspective of targeted CAZyme regions in nonmodel *Trichoderma* species to identify potential highly degradative regions, which will be useful for improving enzymatic degradation technologies. Here, we provide an in-depth analysis of the target lignocellulolytic genomic regions of novel *Trichoderma* strains.

ACKNOWLEDGMENTS

We thank CPQBA-CBMAI (Multidisciplinary Center for Chemical, Biological, and Agricultural Research, Brazilian Collection of Environment and Industry Microorganisms) for providing the fungal samples and CBMEG (Center for Molecular Biology and Genetic Engineering) for providing high-quality infrastructure to accomplish the studies.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

DATA AVAILABILITY STATEMENT

All the data generated or analyzed in this study are included in this article (and its supplementary information files). SUPPLEMENTARY MATERIAL 1 contains primer sequences and reference genes. SUPPLEMENTARY MATERIAL 2 presents the OrthoFinder comparative analysis results, orthologs, and others. SUPPLEMENTARY MATERIAL 3 shows the functional annotations, differentially expressed genes, and annotated CAZyme clusters for each strain. The assembled genomic regions were deposited in the NCBI database via BankIt and can be accessed under BioProject numbers PRJNA1028979, Biosamples TH0179 SAMN37846728 and TA0020 SAMN37846658.

FUNDING

Financial support for this work was provided by the São Paulo Research Foundation (Fundação de Amparo à Pesquisa do Estado de São Paulo [FAPESP]; process numbers 2015/09202-0 and 2018/19660-4) and the Coordination for the Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior [CAPES]; Computational Biology Program, process number 88882.160095/2013-01). Paulo Henrique Campiteli de Azevedo received a PhD fellowship from CAPES (88887.612254/2021-00), R.R.R. received PhD fellowships from CAPES (88887.482201/2020-00) and FAPESP (2020/13420-1), Maria Augusta Crivelente Horta received a postdoctoral fellowship from FAPESP (2020/10536-9), and A.P.d.S. received a research fellowship from the Brazilian National Council for Technological and Scientific Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico [CNPq]; process number 312777/2018-3). CAPES was responsible for funding P.H.C.[P.H.C.A.]'s scholarship. The PacBio BAC-sequencing system at the Arizona Genomics Institute was provided by FAPESP.

ORCID

Paulo Henrique Campiteli http://orcid.org/0000-0001-9564-9336
Maria Augusta Horta http://orcid.org/0000-0003-2159-5296
Rafaela Rossi Rosolen http://orcid.org/0000-0002-8964-4034
Juliano Sales Mendes http://orcid.org/0000-0001-7249-2292
Carla Cristina da Silva http://orcid.org/0000-0003-4779-0014
Danilo Sforça http://orcid.org/0000-0003-2304-8712
Anete Pereira de Souza http://orcid.org/0000-0003-3831-9829

LITERATURE CITED

Almeida DA, Horta MAC, Filho JAF, Murad NF, de Souza AP. 2021. The synergistic actions of hydrolytic genes reveal the mechanism of *Trichoderma harzianum* for cellulose degradation. J Biotechnol. 334:1–10. doi:10.1016/j.jbiotec.2021.05.001.

Andlar M, Rezić T, Marđetko N, Kracher D, Ludwig R, Šantek B. 2018. Lignocellulose degradation: an overview of fungi and fungal enzymes involved in lignocellulose degradation. Eng Life Sci. 18(11):768–778. doi:10.1002/elsc.201800039.

Baroncelli R, Piaggeschi G, Fiorini L, Bertolini E, Zapparata A, Pè ME, Sarrocco S, Vannacci G. 2015. Draft whole-genome sequence of the biocontrol agent *Trichoderma harzianum* T6776. Genome Announc. 3(3):e00647–15. doi:10.1128/genomeA.00647-15.

Benocci T, Aguilar□Pontes MV, Kun RS, Seiboth B, de Vries RP, Daly P. 2018. ARA1 regulates not only l-arabinose but also d-galactose catabolism in Trichoderma reesei. FEBS Lett. 592:60–70. doi:10.1002/1873-3468.12932.

Bischof RH, Ramoni J, Seiboth B. 2016. Cellulases and beyond: the first 70 years of the enzyme producer *Trichoderma reesei*. Microb Cell Fact. 15(1):106. doi:10.1186/s12934-016-0507-6.

Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 49(W1):W29–W35. doi:10.1093/nar/gkab335.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Crucello A, Sforça DA, Horta MA, Dos Santos CA, Viana AJ, Beloti LL, de Toledo MA, Vincentz M, Kuroshu RM, de Souza AP. 2015. Analysis of genomic regions of *Trichoderma harzianum* IOC-3844 related to biomass degradation. PLOS ONE. 10(4):e0122122. doi:10.1371/journal.pone.0122122.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29 (1):15–21. doi:10.1093/bioinformatics/bts635.

Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 50(D1): D571–D577. doi:10.1093/nar/gkab1045.

Druzhinina IS, Chenthamara K, Zhang J, Atanasova L, Yang D, Miao Y, Rahimi MJ, Grujic M, Cai F, Pourmehdi S, et al. 2018. Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus *Trichoderma* from its plant-associated hosts. PLOS Genet. 14(4):e1007322. doi:10.1371/journal.pgen.1007322.

Druzhinina IS, Kubicek CP, Komoń-Zelazowska M, Mulaw TB, Bissett J. 2010. The *Trichoderma harzianum* demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages. BMC Evol Biol. 10(1):94. doi:10.1186/1471-2148-10-94.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20 (1):238. doi:10.1186/s13059-019-1832-y.

Filho JAF, Horta MAC, Beloti LL, Dos Santos CA, de Souza AP. 2017. Carbohydrate-active enzymes in *Trichoderma harzianum*: a bioinformatic analysis bioprospecting for key enzymes for the biofuels industry. BMC Genom. 18(1):779. doi:10.1186/s12864-017-4181-9.

Filho JAF, Horta MAC, Dos Santos CA, Almeida DA, Murad NF, Mendes JS, Sforça DA, Silva CBC, Crucello A, de Souza AP. 2020. Integrative genomic analysis of the bioprospection of regulators and accessory enzymes associated with cellulose degradation in a filamentous fungus (*Trichoderma harzianum*). BMC Genom. 21(1):757. doi:10.1186/s12864-020-07158-w.

Filiatrault-Chastel C, Heiss-Blanquet S, Margeot A, Berrin J. 2021. From fungal secretomes to enzymes cocktails: the path forward to bioeconomy. Biotechnol. Adv 52 (107833):107833. doi:10.1016/j.biotechadv.2021.107833.

Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma L-J, Smirnov S, Purcell S, et al. 2003. The genome sequence of the filamentous fungus neurospora crassa. Nature. 422:859–868. doi:10.1038/nature01554.

Galaxy Community. 2022. The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res. 50(W1):W345–W351. doi:10.1093/nar/gkac247

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, et al. 1996. Life with 6000 genes. Science. 274 (546):547–563. doi:10.1126/science.274.5287.546.

Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, et al. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42(D1):D699–D704. doi:10.1093/nar/gkt1183.

Guzmán-Guzmán P, Kumar A, de Los Santos-Villalobos S, Parra-Cota FI, Orozco-Mosqueda MDC, Fadiji AE, Hyder S, Babalola OO, Santoyo G. 2023. *Trichoderma* species: our best fungal allies in the biocontrol of plant diseases-a review. Plants (Basel). 12(3):432. doi:10.3390/plants12030432.

Hewedy OA, Abdel-Lateif KS, Bakr RA. 2020. Genetic diversity and biocontrol efficacy of indigenous *Trichoderma* isolates against Fusarium wilt of pepper. J Bas Microbiol. 60(2):126–135. doi:10.1002/jobm.201900493.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinform. 12 (1):491. doi:10.1186/1471-2105-12-491.

Horta MAC, Filho JAF, Murad NF, de Oliveira Santos E, Dos Santos CA, Mendes JS, Brandão MM, Azzoni SF, de Souza AP. 2018. Network of proteins, enzymes and genes linked to biomass degradation shared by *Trichoderma* species. Sci Rep. 8(1):1341. doi:10.1038/s41598-018-19671-w.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314. doi:10.1093/nar/gky1085.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30(9):1236–1240. doi:10.1093/bioinformatics/btu031.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28(1):27–30. doi:10.1093/nar/28.1.27.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44(D1):D457–D462. doi:10.1093/nar/gkv1070.

Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. PMID: 7463489. 1980;16(2):111–120. 10.1007/BF01731581.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. Genome Res. 27(5):722–736. doi:10.1101/gr.215087.116.

Kubicek CP, Herrera-Estrella A, Seidl-Seiboth V, Martinez DA, Druzhinina IS, Thon M, Zeilinger S, Casas-Flores S, Horwitz BA, Mukherjee PK, et al. 2011. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of *Trichoderma*. Genome Biol. 12(4):R40. doi:10.1186/gb-2011-12-4-r40.

Kubicek CP, Steindorff AS, Chenthamara K, Manganiello G, Henrissat B, Zhang J, Cai F, Kopchinskiy AG, Kubicek EM, Kuo A, et al. 2019. Evolution and comparative genomics of the most common *Trichoderma* species. BMC Genom. 20 (1):485. doi:10.1186/s12864-019-5680-7.

Kun RS, Gomes ACS, Hildén KS, Cerezo SS, R MM, Vries RP. 2019. Developments and opportunities in fungal strain engineering for the production of novel enzymes and enzyme cocktails for plant biomass degradation. Biotechnol. Adv 37(6):107361. doi:10.1016/j.biotechadv.2019.02.017.

Li WC, Huang CH, Chen CL, Chuang YC, Tung SY, Wang TF. 2017. *Trichoderma reesei* complete genome

sequence, repeat-induced point mutation, and partitioning of CAZyme gene clusters. Biotechnol. Biofuels. 10(1):170. doi:10.1186/s13068-017-0825-x.

Li WC, Wang TF. 2021. PacBio long-read sequencing, assembly, and funannotate reannotation of the complete genome of *Trichoderma reesei* QM6a. In: Mach-Aigner AR, Martzy R, editors. *Trichoderma reesei*: methods and Protocols. New York (NY): Springer US; p. 311–329.

Lima PC, Karimian P, Johnston E, Hartley CJ. 2024. The use of *Trichoderma* spp. for the bioconversion of agro-industrial waste biomass via fermentation: a review. Fermentation. 10(9):442. doi:10.3390/fermentation10090442.

Mancini MC, Cardoso-Silva CB, Sforça DA, de Souza AP. 2018. "Targeted sequencing by gene synteny," a new strategy for polyploid species: sequencing and physical structure of a complex sugarcane region. Front Plant Sci. 9:397. doi:10.3389/fpls.2018.00397.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. PLOS Comput Biol. 14(1):e1005944. doi:10.1371/journal.pcbi.1005944.

Mukherjee PK, Horwitz BA, Herrera-Estrella A, Schmoll M, Kenerley CM. 2013. *Trichoderma* Research in the Genome Era. Annu Rev Phytopathol. 51(1):105–109. doi:10.1146/annurev-phyto-082712-102353.

Nascimento VC, Rodrigues-Santos KC, Carvalho-Alencar KL, Castro MB, Kruger RH, Lopes FAC. 2022. *Trichoderma*: biological control efficiency and perspectives for the Brazilian midwest states and Tocantins. Braz J Biol. 82: e260161. doi:10.1590/1519-6984.260161.

Ohm RA, Riley R, Salamov A, Min B, Choi IG, Grigoriev IV. 2014. Genomics of wood-degrading fungi. Fungal Genet Biol. 72:82–90. doi:10.1016/j.fgb.2014.05.001.

Pachauri S, Zaid R, Sherkhane PD, Easa J, Viterbo A, Chet I, Horwitz BA, Mukherjee PK. 2023. Comparative phenotypic, genomic, and transcriptomic analyses of two contrasting strains of the plant beneficial fungus *Trichoderma virens*. Microbiol Spectr. 11:e0302422. doi:10.1128/spectrum.03024-22.

Palmer MJ, Stajich J. Funannotate v1.8.1: eukaryotic Genome Annotation, 2020; version 1.8.1 [Software]. https://zenodo.org/records/4054262.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 33(3):290–295. doi:10.1038/nbt.3122.

Peterson DG, Tomkins JP, Frisch DA, Wing RA, Paterson AH. 2000. Construction of plant bacterial artificial chromosome (BAC) libraries: an illustrated guide. J Agric Genom. 5:1–100. doi:10.1007/978-94-011-4217-5_1.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 33(Web Server):W116–W120. doi:10.1093/nar/gki442.

Raeder U, Broda P. 1985. Rapid preparation of DNA from filamentous fungi. Lett Appl Microbiol. 1(1):17–20. doi:10.1111/j.1472-765X.1985.tb01479.x.

Raulo R, Heuson E, Froidevaux, Froidevaux R, Phalip V. 2021. Combining analytical approaches for better lignocellulosic biomass degradation: a way of improving fungal enzymatic cocktails? Biotechnol Lett. 43(12):2283–2298. doi:10.1007/s10529-021-03201-2.

Riley R, Salamov AA, Brown DW, Nagy LG, Floudas D, Held BW, Levasseur A, Lombard V, Morin E, Otillar R, et al. 2014. Extensive sampling of basidiomycete genomes demonstrates inadequacy of the white-rot/brown-rot paradigm for wood decay fungi. Proc Natl Acad Sci U S A. 111 (27):9923–9928. doi:10.1073/pnas.1400592111.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edge: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26 (1):139–140. doi:10.1093/bioinformatics/btp616.

Rokas A, Wisecaver JH, Lind AL. 2018. The birth, evolution and death of metabolic gene clusters in fungi. Nat Rev Microbiol. 16(12):731–744. doi:10.1038/s41579-018-0075-3.

Rosolen RR, Aono AH, Almeida DA, Filho JAF, Horta MAC, De Souza AP. 2022. Network analysis reveals different cellulose degradation strategies across *Trichoderma harzianum* strains associated with XYR1 and CRE1. Front Genet. 13:807243. doi:10.3389/fgene.2022.807243.

Rosolen RR, Horta MAC, de Azevedo PHC, da Silva CC, Sforca DA, Goldman GH, de Souza AP. 2023. Whole-genome sequencing and comparative genomic analysis of potential biotechnological strains of *Trichoderma harzianum*, Trichoderma atroviride, and *Trichoderma reesei*. *Trichoderma Atroviride*, and *Trichoderma Reesei*. Mol. Genet Genom. 298 (3):735–754. doi:10.1007/s00438-023-02013-5.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4(4):406–425. doi: 10.1093/oxfordjournals.molbev.a040454. PMID: 3447015.

Salazar-Cerezo S, de Vries Rp, Garrigues S, de Vries RP. 2023. Strategies for the Development of Industrial Fungal Producing Strains, J Fungi (Basel). 8; 9(8): 834, 10.3390/jof9080834. PMID: 37623605; PMCID: PMC10455633.

Sardi M, Gasch AP. 2017. Incorporating comparative genomics into the design–test–learn cycle of microbial strain engineering. FEMS Yeast Research. 17(5):42. doi:10.1093/femsyr/fox042.

Schalamun M, Schmoll M. 2022. *Trichoderma* – genomes and genomics as treasure troves for research towards biology, biotechnology and agriculture. Front Fungal Biol. 3:1002161. doi:10.3389/ffunb.2022.1002161.

Slot JC. 2017. Fungal gene cluster diversity and evolution. Adv Genet. 100:141–178. doi:10.1016/bs.adgen.2017.09.005.

Sood M, Kapoor D, Kumar V, Sheteiwy MS, Ramakrishnan M, Landi M, Araniti F, Sharma A. 2020. *Trichoderma: the "Secrets" of a Multitalented Biocontrol Agent*. Plants (Basel). 9(6):762. doi:10.3390/plants9060762.

Srivastava M, Shahid M. 2014. *Trichoderma* genome to genomics: a review. J Data Min Genom Proteom. 5:1000162. doi:10.4172/2153-0602.1000162.

Stracquadanio C, Quiles JM, Meca G, Cacciola SO. 2020. Antifungal activity of bioactive metabolites produced by *Trichoderma asperellum* and *Trichoderma atroviride* in liquid medium. J Fungi Basel. 6(4):263. doi:10.3390/jof6040263.

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol Aug. Epub 2007 May 7. PMID: 17488738. 2007;24(8):1596–1599. 10.1093/molbev/msm092.

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, Nov 11; 22(22): 4673–4680, 10.1093/nar/22.22.4673. PMID: 7984417; PMCID: PMC308517.

Trivedi UH, Cézard T, Bridgett S, Montazam A, Nichols J, Blaxter M, Gharbi K. 2014. Quality control of next-generation sequencing data without a reference. Front Genet. 5:111. doi:10.3389/fgene.2014.00111.

Tyśkiewicz R, Nowak A, Ozimek E, Jaroszuk-Ściseł J. 2022. *Trichoderma*: the current status of its application in agriculture for the biocontrol of fungal phytopathogens and stimulation of plant growth. Int J Mol Sci. 23(4):2329. doi:10.3390/ijms23042329.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res. 40(15):e115. doi:10.1093/nar/gks596.

Visendi P, Berkman PJ, Hayashi S, Golicz AA, Bayer PE, Ruperao P, Hurgobin B, Montenegro J, Chan CK, Staňková H, et al. 2016. An efficient approach to BAC based assembly of complex genomes. Plant Methods. 12(1):2. doi:10.1186/s13007-016-0107-9.

Waterston RH, Lander ES, Sulston JE. 2002. On the sequencing of the human genome. Proc Natl Acad Sci U S A. 99(6):3712–3716. doi:10.1073/pnas.042692499.

Wisecaver JH, Rokas A. 2015. Fungal metabolic gene clusters-caravans traveling across genomes and environments. Front Microbiol. 6:161. doi:10.3389/fmicb.2015.00161.

Xu L, Gonzalez Ramos L, Lyra VM, Wiebenga C, Grigoriev AD, Vries IG, Miia RP, Mäkelä R, Mao MR, Peng M. 2024. Genome-wide prediction and transcriptome analysis of sugar transporters in four ascomycete fungi. Bioresour Technol. 391:130006. doi:10.1016/j.biortech.2023.130006.

Xu M, Wang Y, Zhao Z, Gao G, Huang SX, Kang Q, He X, Lin S, Pang X, Deng Z, et al. 2016. Functional genome mining for metabolites encoded by large gene clusters through heterologous expression of a whole-genome bacterial artificial chromosome library in streptomyces spp. Appl Environ Microbiol. 82(19):5795–5805. doi:10.1128/aem.01383-16.

Zeilinger S, Gruber S, Bansal R, Mukherjee PK. 2016. Secondary metabolism in *Trichoderma* – chemistry meets genomics. Fungal Biol Rev. 30(2):74–90. doi:10.1016/j.fbr.2016.05.001.

Zhang Z, Running KLD, Seneviratne S, Haugrud ARP, Szabo-Hever A, Shi G, Brueggeman R, Xu SS, Friesen TL, Faris JD. 2021. A protein kinase-major sperm protein gene hijacked by a necrotrophic fungal pathogen triggers disease susceptibility in wheat. Plant J. 106(3):720–732. doi:10.1111/tpj.15194.

Zheng J, Hu B, Zhang X, Ge Q, Yan Y, Akresi J, Piyush V, Huang L, Yin Y. 2023. dbCAN-seq update: cAZyme gene clusters and substrates in microbiomes. Nucleic Acids Res. 51(D1):D557–D563. doi:10.1093/nar/gkac1068.