Whole-Genome Sequencing and Comparative Genomic Analysis of Potential Biotechnological Strains from *Trichoderma harzianum*, *Trichoderma atroviride*, and *Trichoderma reesei*

Rafaela Rossi Rosolen[1,2], Maria Augusta Crivelente Horta[1,4], Paulo Henrique Campiteli de Azevedo[1,2], Carla Cristina da Silva[1], Danilo Augusto Sforca[1], Gustavo Henrique Goldman[4], and Anete Pereira de Souza[1,3*]

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

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

[3]Department of Plant Biology, Institute of Biology, UNICAMP, Cidade Universitária Zeferino Vaz, Rua Monteiro Lobato, Campinas, SP, Brazil

[4]Faculty of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo (USP), Ribeirão Preto, SP, Brazil

*** Correspondence:**
Anete Pereira de Souza
anete@unicamp.br

Number of words: 5644

Number of figures: 7

Number of tables: 3

Keywords: *Trichoderma*, genomes, comparative analysis, orthology analysis, structural variant analysis

Abstract

Fungi of the genus *Trichoderma* exhibit high genetic diversity and can thus be utilized in a large range of biotechnological applications. While *Trichoderma reesei* is the primary source of industrial enzymatic cocktails, *Trichoderma atroviride* and *Trichoderma harzianum* are widely used as commercial biocontrol agents against plant diseases. Recently, *T. harzianum* IOC-3844 (Th3844) and *T. harzianum* CBMAI-0179 (Th0179) demonstrated great potential in the enzymatic conversion of lignocellulose into fermentable sugars. Despite such potential, the genomes of both hydrolytic strains remain unclear. Herein, we performed whole-genome sequencing and assembly of Th3844 and Th0179 strains. To assess the genetic diversity within the genus *Trichoderma*, the results of both strains were compared with those from *T. atroviride* CBMAI-00020 (Ta0020) and *T. reesei* CBMAI-0711 (Tr0711). The resulting assembly revealed a total length of 40 Mb, (Th3844), 39 Mb (Th0179), 36 Mb (Ta0020), and 32 Mb (Tr0711), in which 10,786 (Th3844), 11,322 (Th0179), 10,082 (Ta0020), and 8,796 (Tr0711) genes were predicted. Then, the annotation of the predicted CDS sequences from the evaluated strain genes revealed 413 (Th3844), 413 (Th0179), 377 (Ta0020), and 329 (Tr0711) CAZymes. The orthology analysis revealed 18,349 orthogroups, which encompassed 95% of the total genes, 3,378 orthogroups with all species present, and 408 species-specific orthogroups. A genome-

wide phylogenetic analysis modeled from the 1,864 single-copy orthogroups provided details on the relationship of the newly sequenced species with other *Trichoderma* and with more evolutionarily distant genera. Structural variants revealed genomic rearrangements between Th3844, Th0179, Ta0020, and Tr0711 with the *T. reesei* QM6a reference genome as well as the functional effects of such variants on the evaluated strains. In conclusion, the findings presented herein allow the genetic diversity of the evaluated strains, including those from the same species, to be viewed, offering opportunities to further explore such fungal genomes in future biotechnological and industrial applications.

1 Introduction

Fungi of the genus *Trichoderma* are characterized by their considerable nutritional versatility (Sharma et al., 2019), which allows them to be employed in a wide range of biotechnological applications (Kidwai and Nehra, 2017). For example, *Trichoderma reesei* is the primary fungal source of the industrial cellulases and hemicellulases present in enzymatic cocktails (Bischof et al., 2016). In addition to enzymatic activities, the capacity of biocontrol against plant pathogenic fungi has been widely explored in *Trichoderma harzianum* and *Trichoderma atroviride* (Medeiros et al., 2017; Saravanakumar et al., 2017). Recently, *T. harzianum* strains were explored for their enzymatic potential and were demonstrated to be useful for improving the lignocellulosic conversion into sugars during second-generation ethanol (2G ethanol) production (Almeida et al., 2021; Delabona et al., 2020a; Motta et al., 2021; Zhang et al., 2020a).

The phenotypic and ecological heterogeneity across fungi of the same genera is reflected, in part, by the diversity observed within their genomes (Priest et al., 2020). In this way, the diverse and important roles of fungi, as well as the technological advances in next-generation sequencing, have motivated broad efforts to sequence several fungal genomes (Ganesh Kumar et al., 2021; Hagestad et al., 2021; Nagel et al., 2021; Varga et al., 2019; Wu et al., 2018). Since the genome of *T. reesei* QM6a was first presented (Martinez et al., 2008), *Trichoderma* sequencing studies have increased with the goal of better understanding the biological and ecological roles of *Trichoderma* to improve their applications (Druzhinina et al., 2018b; Horta et al., 2014; Kubicek et al., 2011; Kubicek et al., 2019; Li et al., 2017; Schmoll et al., 2016). Such a growing number of sequenced species may help reveal the molecular basis for the specific features of diverse *Trichoderma* strains.

Previously, the transcriptional profiles of two *T. harzianum* strains, *T. harzianum* IOC-3844 (Th3844) and *T. harzianum* CBMAI-0179 (Th0179), were analyzed under cellulose degradation conditions and compared with those from *T. atroviride* CBMAI-0020 (Ta0020) and *T. reesei* CBMAI-0711 (Tr0711) (Almeida et al., 2021; Horta et al., 2018). Such studies have suggested the great potential of both *T. harzianum* strains as hydrolytic enzyme producers, and this was similar to Tr0711, while Ta0020 showed a low cellulolytic ability. Furthermore, differences in the transcription regulation within hydrolytic enzyme expression were observed for Th3844 and Th0179 (Rosolen et al., 2021), highlighting the genetic differences between such strains. Although previous studies investigated the Th3844 genomic regions, which are related to biomass degradation, through bacterial artificial chromosome (BAC) library construction (Crucello et al., 2015; Ferreira Filho et al., 2017), genomic information regarding the hydrolytic strains of *T. harzianum*, Th3844 and Th0179 remains unclear.

In this study, Pacific Biosciences (PacBio) (Ardui et al., 2018) technology was used to obtain highly contiguous de novo assemblies and to describe the genetic variation present among Th3844 and Th0179. Aiming to expand knowledge on the genetic diversity within the genus *Trichoderma*, the results obtained for *T. harzianum* strains were compared to those from *T. atroviride* and *T. reesei*. The

chosen species are appropriate for the study's goal because while *T. atroviride* is a biocontrol species that is distantly related to the lignocellulolytic species *T. reesei* (Druzhinina et al., 2006), representing a well-defined phylogenetic species (Dodd et al., 2003), *T. harzianum sensu lato* is also commonly used in biocontrol but constitutes a complex of several cryptic species (Chaverri et al., 2015; Druzhinina et al., 2010).

After performing whole-genome annotation, we investigated the content of carbohydrate-active enzymes (CAZymes) (Cantarel et al., 2009) that were distributed among the studied genomes. To thoroughly investigate the genetic variability across the four evaluated strains, we explored the structural variants (SVs), which represent a major form of genetic and phenotypic variations that are inherited and polymorphic in species (Mills et al., 2011), between them and *T. reesei* QM6a, the reference genome (Martinez et al., 2008). In addition, by performing a comparative genomic analysis across the genus *Trichoderma* and more evolutionarily distant genera, the orthologs and the orthogroups across them were identified, and the rooted gene tree based on the single-copy orthologs was inferred.

The genomic resources we provide herein significantly extend our knowledge regarding the evolution and basic biology of the evaluated strains, and this may increase their biotechnological employment. The results from this study might also increase the availability of genomic data, which can be used to perform comparative studies to correlate phenotypic differences in the genetic diversity of *Trichoderma* species; therefore, the study may help to improve the search for enzymes with enhanced properties and provide aid toward improving the production of chemicals and enzymes in such fungi.

2 Material and methods

2.1 Fungal strains and culture conditions

The species originated from the Brazilian Collection of Environment and Industry Microorganisms (CBMAI), which is located in the Chemical, Biological, and Agricultural Pluridisciplinary Research Center (CPQBA) at the University of Campinas (UNICAMP), Brazil. The identity of *Trichoderma* isolates was authenticated by CBMAI based on phylogenetic studies of their internal transcribed spacer (ITS) region and translational elongation factor 1 (*tef1*) marker gene. Briefly, Th3844, Th0179, Ta0020, and Tr0711 strains were cultivated on potato dextrose agar (PDA) solid medium (ampicillin 100 µg/ml and chloramphenicol 34 µg/ml) for 3 days at 28 °C. Conidia were harvested, and an initial spore solution was used to inoculate 500 mL of potato dextrose broth (PDB) medium. The cultivation process was performed in biological triplicates for 72 h at 28 °C and 200 rpm for all evaluated strains. Then, mycelial samples were harvested using Miracloth (Millipore), frozen using liquid nitrogen, and stored at -80 °C. Frozen material was used for DNA extraction.

2.2 DNA extraction and sequencing

The ground fungal tissue was suspended using lysis buffer, then phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma, US) was added. After centrifugation at 4 °C and 13,000 rpm for 10 min, the aqueous layer was collected, and genomic DNA was precipitated via the addition of isopropanol. DNA was harvested by centrifugation at 4 °C and 13,000 rpm for 10 min, and the pellet was washed with 70% ethanol, followed by centrifugation at 4 °C and 13,000 rpm for 5 min. After a second washing with 95% ethanol and centrifugation at 4 °C and 13,000 rpm for 5 min, the pellet was dried at room temperature and dissolved in TE buffer. Before the quality control steps, the DNA was subjected to RNAse treatment.

The quantity of the extracted gDNA was determined by measuring the absorbance at 260 nm using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and Qubit Fluorometer (Thermo Fisher Scientific). The quality of extracted gDNA was assessed through 0.8% agarose gel electrophoresis. HiFi sequencing libraries were prepared according to the PacBio protocol, and sequencing was performed at the Arizona Genomics Institute (AGI; Tucson, USA) using a SMRT DNA sequencing system from PacBio (PacBio RSII platform).

2.3 Genome assembly

The data were transferred to a local server, and the genomes were assembled de novo using Canu software (v.2.1) (-pacbio – hifi, and a genome estimate equal to 40 Mb for all evaluated strains), which was developed for long-read sequencing (Koren et al., 2017). Genome integrity was assessed using the Quality Assessment Tool (QUAST) (Gurevich et al., 2013) (v.5.0.2) and Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al., 2015) (v.4.1.4) tools. The Nucmer alignment tool from the MUMmer (v.4.0.0beta2) toolbox (Kurtz et al., 2004; Marçais et al., 2018) was used to perform the whole-genome alignments between the evaluated strains.

2.4 Gene prediction and functional annotation

Gene prediction was performed using AUGUSTUS (v.3.3.3) (Stanke et al., 2006) through gene models, which were built from *T. harzianum* T6776, *T. atroviride* IMI206040, and *T. reesei* QM6a (TrainAugustus (v.3.3.3)), together with the MAKER (Cantarel et al., 2008) (v.2.31.11). Such programs are implemented on the Galaxy platform. The predicted genes were functionally annotated by searching for homologous sequences in the UniProt (The UniProt, 2021), eggNOG-mapper v.2 (Cantalapiedra et al., 2021), and Protein Annotation with Z score (PANNZER2) (Törönen et al., 2018) databases. Transmembrane proteins were predicted using TMHMM v.2.0 (Krogh et al., 2001). For the annotation of CAZymes, we used CDS sequences as homology search queries against the database of the dbCAN2 server (Zhang et al., 2018), which integrates (I) DIAMOND (E-Value < 1e-102) (Buchfink et al., 2015), (II) HMMER (E-Value < 1e-15, coverage > 0.35) (Finn et al., 2011), and Hotpep (Frequency > 2.6, Hits > 6) (Busk et al., 2017) tools. We considered all CDSs as true hits if they were predicted by at least two tools. Coverages were estimated with QualiMap (Okonechnikov et al., 2016) (v.2.2.2c) using minimap2 (Li, 2018) v. 2.17 + galaxy4, which were both implemented on the Galaxy platform (Afgan et al., 2018).

2.5 Ortholog identification and clustering

The proteomes of Th3844, Th0179, Ta0020, and Tr0711 were compared with the *Trichoderma* spp. proteomes that are available on NCBI databases. *Fusarium* spp., *Aspergillus* spp., and *Neurospora* spp. were used as outgroup. For this analysis, we used the software OrthoFinder (Emms and Kelly, 2015, 2019) v2.5.2, which clustered the protein sequences of fungi into orthologous groups and allowed the phylogenetic relationships between them to be identified. The consensus species tree was inferred using STAG algorithm (Emms and Kelly, 2018) and rooted using STRIDE algorithm (Emms and Kelly, 2017), which are implemented on the OrthoFinder program. The resulting tree from the OrthoFinder analysis was visualized and edited using Interactive Tree of Life (iTOL) v6 (Letunic and Bork, 2007).

2.6 Long-read structural variant analysis

SVs were identified by aligning the PacBio HiFi reads from Th3844, Th0179, Ta0020, and Tr0711 with the *T. reesei* QM6a reference genome (Martinez et al., 2008) using the software Map with

BWA-MEM (Li and Durbin, 2010) v.0.7.17.2 with (-x pacbio, sort by chromosomal coordinates). The duplicate reads in the BAM file were identified and marked using the tool MarkDuplicates (Institute) v.2.18.2.2. Variants were called using Sniffles (Sedlazeck et al., 2018) v.1.0.12+galaxy0 allowing for a minimum support of 10 (--min_support), maximum number of splits of 7 (--max_num_splits), maximum distance of 1000 (--max_distance), minimum length of 30 (--min_length), minimum mapping quality of 20 (--minmapping_qual), and CCS reads option (--ccs_reads). SVs were annotated using SnpEff (v.4.3+T. galaxy1) (Cingolani et al., 2012), which allowed the effects of variants in genome sequences to be categorized. Such tools are implemented on the Galaxy platform (Afgan et al., 2018).

3 Results

3.1 Strain cultivations and evaluation of extracted DNA

First, we cultivated Th3844, Th0179, Ta0020, and Tr0711, which can be identified by common morphological characteristics, such as a bright green conidial pigment and a repetitive branch (Figure 1). Next, the DNA from the evaluated *Trichoderma* isolates was extracted, and its integrity and quality were assessed (Supplementary Material 1: Supplementary Table 1 and Supplementary Material 1: Supplementary Figure 1).

3.2 Assembled genomic features and general comparison across *Trichoderma* spp.

In the present study, we introduced the whole-genome sequences of Th3844, Th0179, Ta0020, and Tr0711 (Table 1). Overall, the genomes of the evaluated *Trichoderma* spp. varied in a number of contigs (14–26), sizes (32–40 Mb) and gene contents (8,796–11,322 genes). In comparison with the other strains, Th0711 contains the smallest gene repertoire, while Th3844 contains the highest gene repertoire. To assess the completeness and integrity of the assembled genomes, BUSCO analysis was performed. For all evaluated strains, over 90% of genes were complete. Although the genome of Th3844 presented a considerable number of missing genes (9.7%), its assembled genome exhibited a lower degree of fragmentation compared to that of other genomes.

Considering the genome sequencing coverage, genome size, GC content, length metrics (N50 and L50 values), assembly level, and number of genes, the genomes of Th3844, Th0179, Ta0020, and Tr0711 were compared to other fungal genome references (Baroncelli et al., 2015; Chung et al., 2021; Kubicek et al., 2011; Li et al., 2017; Martinez et al., 2008) (Table 2). All *T. harzianum* genomes were similar in size and GC content. The same profile was observed for the *T. atroviride* and *T. reesei* genomes. Large differences were found for the genome sequencing coverage, in which Th3844, Th0179, Ta0020, and Tr0711 presented higher values than those reported for other strains in the literature (Baroncelli et al.; Kubicek et al., 2011; Li et al., 2017). In regard to quality, except for Tr0711, the genomes assembled in this study showed a lower degree of fragmentation compared to those previously available (Baroncelli et al., 2015; Chung et al., 2021; Kubicek et al., 2011; Li et al., 2017; Martinez et al., 2008).

We also performed alignment analyses to evaluate the similarities and variations between the genomes of the studied strains (I) with *T. reesei* QM6a, which is a model organism for the lignocellulose deconstruction system and has a genome that is assembled at the chromosomic level (Li et al., 2017) (Figure 2), and (II) with their respective reference genomes (Supplementary Material 1: Supplementary Figure 2). The profile of alignment across the genomes illustrated the degree of divergence across the studied strains with *T. reesei* QM6a and with the closest related strain of each evaluated fungus. For each evaluated strain, we observed alignment in different regions of the *T. reesei* QM6a reference

genome. Using *T. reesei* QM6a as a reference genome, we found a total of (I) 12% aligned bases for Th3844, (II) 13% aligned bases for Th0179, (III) 8% aligned bases for Ta0020, and (IV) 95% aligned bases for Tr0711. The total percentage of aligned bases for Th3844 and Th0179 genomes with the *T. harzianum* T6776 genome was approximately 83.5% and 89%, respectively. In addition, for the Ta0020 genome, the total percentage of aligned bases to the *T. atroviride* IMI206040 reference genome was approximately 88%.

3.3 Genome functional annotation and CAZyme distribution

After performing the genome assembly and gene prediction steps, functional annotation was accomplished by a homology search. The functional category distribution regarding the clusters of orthologous groups of proteins (COG) (Tatusov et al., 2000) is shown in Figure 3 and Supplementary Material 1: Supplementary Table 2. Disregarding the (S) function unknown category, the top 5 functional categories were (G) carbohydrate metabolism and transport, (O) posttranslational modification, protein turnover, chaperone functions, (Q) secondary structure, (E) amino acid transport and metabolism, and (U) intracellular trafficking, secretion, and vesicular transport. Overall, the evaluated strains seem to have similar COG assignment profiles. The complete functional annotation of the four genomes is available in Supplementary Material 2: Supplementary Table 3.

Within filamentous fungi, the genus *Trichoderma* is a model system for the production of CAZymes, which includes glycoside hydrolases (GHs), carbohydrate esterases (CEs), glycosyltransferases (GTs), polysaccharide lyases (PLs), auxiliary activities (AAs), and carbohydrate-binding modules (CBMs) (Cantarel et al., 2009). Due to the important role of CAZymes in microparasitism and saprophytic degradation of debris, a thorough investigation regarding the enzyme content in the genomes of the four *Trichoderma* spp. was performed (Supplementary Material 3: Supplementary Table 4). Overall, the genes encoding the CAZymes encompassed approximately 3.8% of the genomes assembled herein. Moreover, to detect similarities and differences between the strains, their CAZyme profiles were compared (Figure 4A).

Among the main CAZyme classes detected in all strains, GHs were overrepresented, and Th3844 (256) and Th0179 (252) had the highest number, followed by Ta0020 (230) and Tr0711 (184). This CAZyme class was followed by GTs as follows: (I) 88 (Th3844), (II) 90 (Th0179), (III) 90 (Ta0020), and (IV) 86 (Tr0711). However, when the CAZymes with predicted signal peptides were analyzed, we observed that only a few GTs were secreted, as follows: (I) 8 (Th3844), (II) 9 (Th0179), (III) 7 (Ta0020), and (IV) 4 (Tr0711) (Figure 4B). The distribution of the different CAZyme families among strains was investigated and is available in Supplementary Material 1: Supplementary Figure 3. For simplification purposes, only the CAZyme families related to biomass degradation are exhibited in Figure 5. Overall, CAZymes from the GH5, AA1, AA3, GH2, and GH3 families were well represented for all strains, and the highest numbers were in Th3844 and Th0179.

Due to the mycoparasitic characteristics of some fungi from the genus *Trichoderma*, CAZymes involved in such activity were also identified in the four evaluated genomes (Supplementary Material 1: Supplementary Figure 3). For example, CAZymes from the GH18 family, which are related to chitin degradation, were present in all evaluated strains, and CAZymes from GH75 are related to chitosan degradation. Additionally, we would like to highlight that other CAZyme classes, including GH16, GH55, and GH64, are related to the mycoparasitic interaction that was present in the genomes of Th3844, Th0179, Ta0020, and Tr0711.

3.4 Orthology analysis, phylogenetic profiling, and structural variant analyses

The predicted proteomes of Th3844, Th0179, Ta0020, and Tr0711 were compared with those of 19 other *Trichoderma* spp. and with those from more phylogenetically distant fungi, including *Fusarium* spp., *Aspergillus* spp., and *Neurospora* spp. (outgroups) (Supplementary Material 4: Supplementary Table 5). We identified a total of 18,349 orthogroups, which encompassed 313,444 genes (95%) in a total of 329,927 genes, i.e., the number of unassigned genes was equal to 16,483 (5%) genes. Moreover, we detected 3,378 orthogroups with all species present, of which 1,864 consisted entirely of single-copy genes and 408 species-specific orthogroups. Fifty percent of all genes were in orthogroups with 29 or more genes (G50 was 29) and were contained in the largest 4,609 orthogroups (O50 was 4609). Regarding the orthologous relationships across the evaluated strains, both *T. harzianum* strains shared the highest number of orthologous genes among them compared to that of the other strains. In relation to the other evaluated strains, Th3844 and Th0179 presented more orthologs that were in common with Ta0020 than with Tr0711 (Table 3). To explore the evolutionary history of Th3844, Th0179, Ta0020, and Tr0711, a rooted species tree was inferred using the 1,864 single-copy orthologous genes that were conserved in the 29 fungi analyzed (Figure 6 and Supplementary Material 1: Supplementary Figure 4). Such phylogenetic analysis indicated that Tr0711 was most closely related to *Trichoderma parareesei* CBS 125925, while Ta0020 was most closely related to *T. atroviride* IMI206040. On the other hand, Th0179 and Th3844 were phylogenetically close to *Trichoderma lentiforme* CFAM-422 and to other *T. harzianum* strains (TR274 and CB226.95), respectively.

The SVs of the evaluated *Trichoderma* isolates were identified by mapping the long reads of the fungi against the reference genome of *T. reesei* QM6a (Martinez et al., 2008). A total of 12,407 (Th3844), 12,763 (Th0179), 11,650 (Ta0020), and 7,103 (Tr0711) SVs were identified for each strain, showing substitution rates of ~1/2,674 nucleotides (Th3844), ~1/2,585 nucleotides (Th0179), ~1/2,832 nucleotides (Ta0020), and ~1/4,655 nucleotides (Tr0711). These SVs included different phenomena that affect gene sequences, such as break ends, deletions, multiple nucleotides and InDels, duplications, insertions, and inversions (Supplementary Material 5: Supplementary Table 6 and Figure 7A). Compared with the other evaluated strains, Tr0711 presented a low number of SVs, while Th0179 displayed the highest number. For all evaluated strains, the most presented SV categories were multiple nucleotides and an InDel, followed by deletions and insertions (Figure 7A). To thoroughly investigate the functional effects of the identified SVs, we performed an annotation of the structural rearrangements, which were placed into different classes based on their predicted effects on protein function. Details of these effects can be found in Supplementary Material 6: Supplementary Table 7, and the most prevalent effects are represented in Figure 7B. For all evaluated strains, the majority of variants presented a modifier impact, which was higher at downstream and upstream genomic locations. Such an effect was more accentuated for both *T. harzianum* strains. In addition, SVs present in transcripts, genes, and intergenic regions were well represented for Ta0020.

4 Discussion

Diversity within the genus *Trichoderma* is evident from the wide range of phenotypes exhibited by the fungi as well as the various ecological roles and industrial purposes the fungi serve (Nakkeeran et al., 2021). Because of their various applications, different *Trichoderma* species have become model organisms for a broad spectrum of physiological phenomena, such as plant cell wall degradation and enzyme production (Fang et al., 2021), biocontrol (Zin and Badaluddin, 2020), and response to light (Schmoll, 2018). Within the genus *Trichoderma*, *T. harzianum* has been used as a commercial biocontrol agent against plant diseases (Fraceto et al., 2018). In addition to their mycoparasitic activities, hydrolytic enzymes from *T. harzianum* strains have demonstrated great potential in the conversion of lignocellulosic biomass into fermentable sugars (Almeida et al., 2021; Delabona et al.,

2020b; Motta et al., 2021; Zhang et al., 2020b). Recently, different types of enzymatic profiles across *Trichoderma* species were reported, in which Th3844 and Th0179 presented a higher hydrolytic potential during growth on cellulose than that of Ta0020 (Almeida et al., 2021); furthermore, differences between Th3844 and Th0179 concerning the transcriptional regulation coordinated by XYR1 and CRE1 during cellulose degradation were reported (Rosolen et al., 2021). Because such diversity in enzyme response might be related to transcriptomic and genomic differences, we aimed to provide foundations for further studies that could investigate such variations at a genomic level.

Technological advances, particularly in long-read sequencing, coupled with the increasing efficiency and decreasing costs across sequencing platforms, enabled fungal genomes to be characterized (Dal Molin et al., 2018; Miyauchi et al., 2020; Wu et al., 2020). Herein, we presented high-quality genome assemblies for two *T. harzianum* strains with hydrolytic potential (Th3844 and Th0179) and, for comparative purposes, for a mycoparasitic species (Ta0020) and saprotrophic species (Tr0711). Thus, *T. reesei* and *T. atroviride* strains were used to assess the genetic differences in the genus *Trichoderma*. With the aim of obtaining high-quality genomes, we employed Canu software for the genome assemblies, and this software has been applied with success in the assembly of fungal genomes (Courtine et al., 2020; Gan et al., 2020; Montoliu-Nerin et al., 2020). Except for Tr0711, the resulting genomics assemblies displayed the highest coverage scores and the lowest fragmentation values compared to those of *Trichoderma virens* Gv29-8 (Kubicek et al., 2011) as well as to *T. harzianum* T6776 (Baroncelli et al., 2015) and *T. atroviride* IMI206040 (Kubicek et al., 2011), which were used as the reference genome in preceding studies from our research group (Almeida et al., 2021; Horta et al., 2018). Although the genomes were not assembled at the chromosome level, the quality of the Th3844, Th0179, Ta0020, and Tr0711 genome assemblies based on the BUSCO value was over 90%. Only the Th3844 genome exhibited 9.7% missing genes. However, even chromosome-level genome assembly does not necessarily achieve a complete BUSCO score (Chung et al., 2021). This may be because the assembly is not 100% accurate, but at the same time, the BUSCO value may not be a perfect indicator to assess genomic qualities. Despite its limitations, without a definitive alternative, BUSCO is still an essential genomic quality assessment tool that includes up-to-date data from many species.

After evaluating the quality of the four assembled genomes, we performed a gene prediction and functional annotation for the datasets. The ecological behavior of the mycoparasites *T. atroviride* and *T. virens*, compared to the plant wall degrader *T. reesei*, is reflected by the sizes of the respective genomes; *T. atroviride* (36 Mb) and *T. virens* (39 Mb) were somewhat larger than the weakly mycoparasitic *T. reesei* (33 Mb) (Kubicek et al., 2011; Martinez et al., 2008). Herein, compared to Th3844 (40 Mb), Th0179 (39 Mb), and Ta0020 (36 Mb), the genome of Tr0711 (32 Mb) was smaller, which might be conceivably as the gene function is lost to mycoparasitism during the evolution of *T. reesei* (Kubicek et al., 2011). In relation to the number of genes, our results showed that Tr0711 presented a smaller gene content than that of Th3844, Th0179, and Ta0020. Such findings corroborated a previous study (Xie et al., 2014), in which 9,143 and 11,865 genes were predicted for *T. reesei* and *T. atroviride*, respectively. In relation to the *T. reesei* QM6a reference genome, the genomes of Th3844, Th0179, Ta0020, and Tr0711 displayed significant structural reorganization, which was more greatly accentuated by an increased phylogenetic distance. Interestingly, this structural reorganization was also observed within strains of the same species, highlighting their genetic diversity.

4.1 Comparative and functional genomics

To obtain insights regarding the functional profile of Th3844, Th0179, Ta0020, and Tr0711, COG analyses of proteins from their genomes were performed. “Carbohydrate metabolism and transport” was a notable COG term for all evaluated strains, suggesting that the genomic arsenal of these fungi is connected to their ability of using carbon sources that are available in the environment. Such characteristics are well known for saprophytic fungal *T. reesei* (Arntzen et al., 2020), and recent studies have observed the same characteristics for *T. harzianum* (Almeida et al., 2021; Delabona et al., 2020a). “Posttranslational modification, protein turnover, chaperone functions” was the second most notable COG term that was present in the four evaluated genomes. Posttranslational modifications (PTMs), which are used by eukaryotic cells to diversify their protein functions and dynamically coordinate their signaling networks, encompass several specific chemical changes that occur on proteins following their synthesis (Ramazi and Zahiri, 2021). The production of intact and functional proteins is a prerequisite for large-scale protein production, and extensive host-specific PTMs often affect the catalytic properties and stability of recombinant enzymes. The high extracellular secretion capability and eukaryotic PTM machinery make *Trichoderma* spp. particularly interesting hosts (Wei et al., 2021). In this context, PTMs are a major factor in the cellulolytic performance of fungal cellulases (Amore et al., 2017; Beckham et al., 2012; Dana et al., 2014), and the impact of plant PTMs on the enzyme performance and stability of the major cellobiohydrolase Cel7A from *T. reesei* has already been determined (van Eerde et al., 2020). In addition, PTMs, especially phosphorylation, of the proteins involved in plant biomass degradation, including CRE1, play an essential role in signal transduction to achieve carbon catabolite repression (CCR) (Han et al., 2020; Horta et al., 2019). Thus, describing this class of *Trichoderma* genomes is essential to understand the impact of alternative PTMs on the catalytic performance and stability of recombinant enzymes. We also would like to highlight that the “secondary structure” and “amino acid transport and metabolism”, which are related to PTMs, were overrepresented COG terms.

To achieve the complete depolymerization of complex lignocellulosic polysaccharides, a repertoire of enzymes that act together on different chemical bonds is needed (Chukwuma et al., 2020). The comparative genomics of *Trichoderma* spp. suggested that mycoparasitic strains, such as *T. vires* and *T. atroviride*, presented a set of genes, including CAZymes and genes encoding secondary metabolites, that were more expressive and related to mycoparasitism compared to those of the saprotrophic species *T. reesei* (Kubicek et al., 2011). Although such fungi are widely used in industry as a source of cellulases and hemicellulases, they have a smaller arsenal of genes that encode the CAZymes related to biomass deconstruction than that of other lignocellulolytic fungi (Martinez et al., 2008). A potential reason for this observation is that its CAZyme content was shaped by loss and a massive horizontal gene transfer (HGT) was gained in enzymes that degrade plant cell walls (Druzhinina et al., 2018a). Herein, compared to the other evaluated strains, Tr0711 also displayed a lower quantity of genes that encode CAZymes. Although the genome of the evaluated strains presented a significant number of GTs, only a few were predicted to be secreted proteins. Such a result could be related to the enzyme activity exhibited by the CAZymes from the GT class, which includes enzymes that are involved in cell wall synthesis in microorganisms and not necessarily in lignocellulose deconstruction (Ardèvol and Rovira, 2015). Furthermore, the presence of putative CAZyme-encoding genes in the genomes of Th0179 and Th3844 provides insight into its lignocellulose-degrading enzyme potential but cannot be directly related to its real degradation ability. In fact, since fungal species rely on different strategies, it has been observed that the number of genes related to the degradation of a given polysaccharide is not necessarily correlated to the extent of its degradation (Arntzen et al., 2020; Kjærbølling et al., 2020). For instance, *T. reesei* relies on the high production levels of a limited set of glycosyl hydrolases (Arntzen et al., 2020). For this reason, CAZy analysis is associated with functional approaches, such as enzymatic activity assays, which provide valuable insight into the actual behavior of the concerned species on specific

lignocellulose substrates. Regarding the CAZyme content, the results found here for the Th3844, Th0179, Ta0020, and Tr0711 genomes follow the same profile as that of a previous study (Fanelli et al., 2018), in which the CAZyme genetic endowment of some strains from *T. harzianum*, including B97 and T6776, was significantly higher than that of *T. atroviride* IMI206040, *T. reesei* QM6a, and *T. virens* Gv-29-8.

In relation to the CAZyme families that are directly associated to the deconstruction of plant biomass, the genomes of Th3844, Th0179, Ta0020, and Tr0711 showed an expressive number of genes encoding GH5, which includes cellulases that are most abundant in fungi (Li and Walton, 2017) and are commonly present in industrial enzymatic cocktails; GH3, which includes β-glucosidases that are frequently secreted into the medium (Guo et al., 2016); and AA3, which is a member of the enzyme arsenal that is auxiliary to GHs (Levasseur et al., 2013). Lytic polysaccharide monooxygenases (LPMOs), which are classified into CAZy auxiliary activity families AA9-AA11 and AA13-AA16, are copper-dependent enzymes that also perform important roles in lignocellulose degradation (Couturier et al., 2018; Monclaro and Filho, 2017). As AA9, AA11, AA13, AA14, and AA16 are exclusive to fungal genomes, multiple genes encoding LPMOs appear to be common in fungal genomes, particularly in Ascomycetes and Basidiomycetes (Kracher et al., 2016). Herein, each of the Th3844, Th0179, Ta0020, and Tr0711 genomes exhibited three AA9 and two AA14 enzymes. Compared to other fungi, the genomes of *Trichoderma* species harbor a high number of chitinolytic genes (Kubicek et al., 2011; Kubicek et al., 2019), reflecting the importance of these enzymes in the mycoparasitic characteristic of fungi. From the *Trichoderma* genomes that were analyzed in detail thus far, the fungal chitinases that belong to the family GH 18 are significantly expanded in *T. virens*, *T. atroviride*, *T. harzianum*, *Trichoderma asperellum*, *Trichoderma gamsii*, and *Trichoderma atrobrunneum* (Fanelli et al., 2018; Kubicek et al., 2011). Similarly, the number of chitosanases (GH75) is enhanced and there are at least five corresponding genes; in contrast, most other fungi have only one or two corresponding genes (Kappel et al., 2020; Kubicek et al., 2011). Furthermore, β-1,3-glucanases that belong to GH families 16, 17, 55, 64, and 81 are expanded in *Trichoderma* mycoparasites compared to other fungi (Fanelli et al., 2018; Kubicek et al., 2011). Here, the CAZyme families that are related to mycoparasitic activity were present in the four genomes studied, and the most were in Th3844 and Th0179.

4.2 Insights into the evolutionary history

Comparative genomics could be a powerful tool for studying fungal evolution and promoting insights into their genetic diversity. In this context, identifying orthology relationships among sequences is an essential step to more thoroughly understand the genetic correlation of particular fungi. Thus, by applying orthology analyses, we could identify orthogroups and orthologs between the evaluated strains, as well as among some other *Trichoderma* spp. and filamentous fungi that are more genetically distant. In the genus *Trichoderma*, several lifestyles have been documented, including saprotrophy, which is a lifestyle that is also observed in other filamentous fungi, such as *Neurospora* spp., *Aspergillus* spp., and *Fusarium* spp. (Arntzen et al., 2020; Corrêa et al., 2020; Najjarzadeh et al., 2021). Thus, the proteomes of some species and strains of such genera were also included in the orthology analysis. Through our results, we may infer that some genus-specific genes are necessary for specific lifestyles and are shared by fungi that have the same lifestyle but are in quite different evolutionary orders. The phylogenetic tree modeled from the orthologous analysis revealed a low bootstrap value for the clade formed by aligning the sequence proteomes from *T. harzianum*. Furthermore, the presence of other *Trichoderma* species, such as *T. lentiforme*, *T. guizhouense*, and *T. simmonsii*, was observed. These observations might be explained by the complex speciation process within the *T. harzianum* species group (Druzhinina et al., 2010); therefore, the phylogenetic

position is uncertain for fungi from these species. However, the molecular identification of Th3844 and Th0179 based on the ITS and *tef1* sequences has already been reported (Rosolen et al., 2021), confirming that both strains were phylogenetically close to other *T. harzianum* strains. Tr0711 was grouped with *T. pararessei* CBS 125925, and both were phylogenetically in proximity to other *T. reesei* strains.

While short reads perform well in the identification of single nucleotide variants (SNVs) and small insertions and deletions (InDels), they are not well suited for detecting changes in larger sequences (Mahmoud et al., 2019). SVs, which include insertions, deletions, duplications, inversions, or translocations that affect ≥ 50 bp (Mills et al., 2011), are more amenable to long-read sequencing (Mitsuhashi and Matsumoto, 2020; Sakamoto et al., 2020). In fungi, SV analysis was employed successfully using reads from third-generation sequencing (Badet et al., 2021; Basile et al., 2021). In this study, we detected SVs by aligning our four genomes against *T. reesei* QM6a (Martinez et al., 2008); although the genomes were assembled at a scaffolding level, unlike the *T. reesei* QM6a reference genome that was assembled at the chromosome level, we chose to proceed with this dataset because its annotation file was available (Li et al., 2017). As expected, due to phylogenetic proximity to the reference genome (Rosolen et al., 2021), Tr0711 showed the lowest number of SVs compared to that of the other strains. However, although *T. atroviride* is phylogenetically distant from *T. reesei* (Rosolen et al., 2021), Ta0020 exhibited fewer SVs than that of both *T. harzianum* strains, and this result might be explained by the uncertain phylogenetic position of fungi in these species (Druzhinina et al., 2010). Although the *T. harzianum* strains are phylogenetically close (Rosolen et al., 2021), genetic variability across the strains is evident when comparing the SVs that were identified from mapping both genomes against *T. reesei* QM6a (Martinez et al., 2008). Such results are consistent with the findings of previous studies in that genetic variations between fungal strains of the same species are not uncommon (Andersen et al., 2011; de Vries et al., 2017; Thanh et al., 2019).

Considering their basic and economic importance, the high-quality genomes found herein might be helpful for better understanding the diversity within the genus *Trichoderma*, as well as improving the biotechnological applications of such fungi. Furthermore, the comparative study of multiple related genomes might be helpful for understanding the evolution of genes that are related to economically important enzymes and for clarifying the evolutionary relationships related to protein function.

Conflicts of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Author contributions

RRR: Writing – original draft, methodology, formal analysis, and visualization. **MACH:** Conceptualization, methodology, and writing – review & editing. **PHCA:** Methodology and resources. **CCS:** Methodology and formal analysis. **DAS:** Methodology and resources. **GHG:** Writing - review & editing. **APS:** Conceptualization, supervision, review & editing, and funding acquisition.

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). RRR received a PhD fellowship from CAPES (88887.482201/2020-00) and FAPESP (2020/13420-1), PHCA received a PhD fellowship from CAPES (88887.612254/2021-00), MACH received a postdoctoral fellowship from FAPESP (2020/10536-9), and APS received a research fellowship from CNPq (312777/2018-3).

Abbreviations

2G ethanol: second-generation ethanol
BAC: bacterial artificial chromosome
BUSCO: benchmarking universal single-copy orthologs
CAZymes: carbohydrate-active enzymes
CBMs: carbohydrate-binding modules
CBMAI: Brazilian collection of environmental and industrial microorganisms
CCR: carbon catabolite repression
CEs: carbohydrate esterases
COG: clusters of orthologous groups of proteins
CPQBA: chemical, biological, and agricultural pluridisciplinary research center
GHs: glycoside hydrolases
GTs: glycosyltransferases
HGT: horizontal gene transfer
InDels: small insertion and deletions
iTOL: interactive tree of life
ITS: internal transcribed spacer
LPMOs: lytic polysaccharide mono-oxygenases
PacBio: pacific biosciences
PANNZER2: protein annotation with z score
PDA: potato dextrose agar
PDB: potato dextrose broth
PLs: polysaccharide lyases
PTMs: posttranslational modifications
QUAST: quality assessment tool
SNVs: single nucleotide variants
SVs: structural variants
Ta0020: *Trichoderma atroviride* CBMAI-0020
tef1: translational elongation factor 1
Th0179: *Trichoderma harzianum* CBMAI-0179
Th3844: *Trichoderma harzianum* IOC-3844

Acknowledgments

We are grateful to CBMAI Campinas and SP for conceiving the fungal isolates used in the current study; the Center of Molecular Biology and Genetic Engineering (CBMEG) at the University of Campinas and SP for the use of the center and laboratory space; and the São Paulo Research Foundation (FAPESP), the Coordination of Improvement of Higher Education Personnel (CAPES, Computational Biology Program), and the Brazilian National Council for Technological and Scientific Development (CNPq) for supporting the project and researchers.

Data availability statement

All data generated or analyzed in this study are included in this published article (and its supplementary information files). The raw datasets and the assembled genomes were deposited in the NCBI Sequence Read Archive and can be accessed under BioProject number PRJNA781962 and BioSample number, as follow SAMN23309297 (Th3844), SAMN23309298 (Th0179), SAMN23309299 (Ta0020), and SAMN23309300 (Tr0711).

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research 46, W537-W544. https://doi.org/10.1093/nar/gky379.

Almeida, D.A., Horta, M.A.C., Ferreira Filho, J.A., Murad, N.F., de Souza, A.P., 2021. The synergistic actions of hydrolytic genes reveal the mechanism of Trichoderma harzianum for cellulose degradation. Journal of Biotechnology 334, 1-10. https://doi.org/https://doi.org/10.1016/j.jbiotec.2021.05.001.

Amore, A., Knott, B.C., Supekar, N.T., Shajahan, A., Azadi, P., Zhao, P., Wells, L., Linger, J.G., Hobdey, S.E., Vander Wall, T.A., Shollenberger, T., Yarbrough, J.M., Tan, Z., Crowley, M.F., Himmel, M.E., Decker, S.R., Beckham, G.T., Taylor, L.E., 2nd, 2017. Distinct roles of N- and O-glycans in cellulase activity and stability. Proceedings of the National Academy of Sciences of the United States of America 114, 13667-13672. https://doi.org/10.1073/pnas.1714249114.

Andersen, M.R., Salazar, M.P., Schaap, P.J., van de Vondervoort, P.J.I., Culley, D., Thykaer, J., Frisvad, J.C., Nielsen, K.F., Albang, R., Albermann, K., Berka, R.M., Braus, G.H., Braus-Stromeyer, S.A., Corrochano, L.M., Dai, Z., van Dijck, P.W.M., Hofmann, G., Lasure, L.L., Magnuson, J.K., Menke, H., Meijer, M., Meijer, S.L., Nielsen, J.B., Nielsen, M.L., van Ooyen, A.J.J., Pel, H.J., Poulsen, L., Samson, R.A., Stam, H., Tsang, A., van den Brink, J.M., Atkins, A., Aerts, A., Shapiro, H., Pangilinan, J., Salamov, A., Lou, Y., Lindquist, E., Lucas, S., Grimwood, J., Grigoriev, I.V., Kubicek, C.P., Martinez, D., van Peij, N.N.M.E., Roubos, J.A., Nielsen, J., Baker, S.E., 2011. Comparative genomics of citric-acid-producing Aspergillus niger ATCC 1015 versus enzyme-producing CBS 513.88. Genome research 21, 885-897. https://doi.org/10.1101/gr.112169.110.

Ardèvol, A., Rovira, C., 2015. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. Journal of the American Chemical Society 137, 7528-7547. https://doi.org/10.1021/jacs.5b01156.

Ardui, S., Ameur, A., Vermeesch, J.R., Hestand, M.S., 2018. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Research 46, 2159-2168. https://doi.org/10.1093/nar/gky066.

562 Arntzen, M.Ø., Bengtsson, O., Várnai, A., Delogu, F., Mathiesen, G., Eijsink, V.G.H., 2020.
563 Quantitative comparison of the biomass-degrading enzyme repertoires of five filamentous
564 fungi. Scientific Reports 10, 20267. https://doi.org/10.1038/s41598-020-75217-z.

565 Badet, T., Fouché, S., Hartmann, F.E., Zala, M., Croll, D., 2021. Machine-learning predicts genomic
566 determinants of meiosis-driven structural variation in a eukaryotic pathogen. Nature
567 Communications 12, 3551. https://doi.org/10.1038/s41467-021-23862-x.

568 Baroncelli, R., Piaggeschi, G., Fiorini, L., Bertolini, E., Zapparata, A., Pè Mario, E., Sarrocco, S.,
569 Vannacci, G., Draft Whole-Genome Sequence of the Biocontrol Agent Trichoderma
570 harzianum T6776. Genome Announcements 3, e00647-00615.
571 https://doi.org/10.1128/genomeA.00647-15.

572 Baroncelli, R., Piaggeschi, G., Fiorini, L., Bertolini, E., Zapparata, A., Pè, M.E., Sarrocco, S.,
573 Vannacci, G., 2015. Draft Whole-Genome Sequence of the Biocontrol Agent Trichoderma
574 harzianum T6776. Genome announcements 3, e00647-00615.
575 https://doi.org/10.1128/genomeA.00647-15.

576 Basile, A., De Pascale, F., Bianca, F., Rossi, A., Frizzarin, M., De Bernardini, N., Bosaro, M.,
577 Baldisseri, A., Antoniali, P., Lopreiato, R., Treu, L., Campanaro, S., 2021. Large-scale
578 sequencing and comparative analysis of oenological Saccharomyces cerevisiae strains
579 supported by nanopore refinement of key genomes. Food Microbiology 97, 103753.
580 https://doi.org/https://doi.org/10.1016/j.fm.2021.103753.

581 Beckham, G.T., Dai, Z., Matthews, J.F., Momany, M., Payne, C.M., Adney, W.S., Baker, S.E.,
582 Himmel, M.E., 2012. Harnessing glycosylation to improve cellulase activity. Current Opinion
583 in Biotechnology 23, 338-345. https://doi.org/https://doi.org/10.1016/j.copbio.2011.11.030.

584 Bischof, R.H., Ramoni, J., Seiboth, B., 2016. Cellulases and beyond: the first 70 years of the enzyme
585 producer Trichoderma reesei. Microb. Cell Fact. 15, 106. https://doi.org/10.1186/s12934-016-
586 0507-6.

587 Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND.
588 Nature Methods 12, 59-60. https://doi.org/10.1038/nmeth.3176.

589 Busk, P.K., Pilgaard, B., Lezyk, M.J., Meyer, A.S., Lange, L., 2017. Homology to peptide pattern for
590 annotation of carbohydrate-active enzymes and prediction of function. BMC bioinformatics
591 18, 214-214. https://doi.org/10.1186/s12859-017-1625-9.

592 Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., 2021. eggNOG-
593 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
594 Metagenomic Scale. bioRxiv 2021.2006.2003.446934.
595 https://doi.org/10.1101/2021.06.03.446934.

596 Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The
597 Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.
598 Nucleic acids research 37, D233-D238. https://doi.org/10.1093/nar/gkn663.

599 Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado,
600 A., Yandell, M., 2008. MAKER: an easy-to-use annotation pipeline designed for emerging
601 model organism genomes. Genome research 18, 188-196. https://doi.org/10.1101/gr.6743907.

602 Chaverri, P., Branco-Rocha, F., Jaklitsch, W., Gazis, R., Degenkolb, T., Samuels, G.J., 2015.
603 Systematics of the Trichoderma harzianum species complex and the re-identification of
604 commercial biocontrol strains. Mycologia 107, 558-590. https://doi.org/10.3852/14-147.

605 Chukwuma, O.B., Rafatullah, M., Tajarudin, H.A., Ismail, N., 2020. Lignocellulolytic Enzymes in
606 Biotechnological and Industrial Processes: A Review. Sustainability 12,
607 https://doi.org/10.3390/su12187282.

608 Chung, D., Kwon, Y.M., Yang, Y., 2021. Telomere-to-telomere genome assembly of asparaginase-
609 producing Trichoderma simmonsii. BMC Genomics 22, 830. https://doi.org/10.1186/s12864-
610 021-08162-4.

611 Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden,
612 D.M., 2012. A program for annotating and predicting the effects of single nucleotide
613 polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-
614 2; iso-3. Fly 6, 80-92. https://doi.org/10.4161/fly.19695.

615 Corrêa, C.L., Midorikawa, G.E.O., Filho, E.X.F., Noronha, E.F., Alves, G.S.C., Togawa, R.C., Silva-
616 Junior, O.B., Costa, M.M.d.C., Grynberg, P., Miller, R.N.G., 2020. Transcriptome Profiling-
617 Based Analysis of Carbohydrate-Active Enzymes in Aspergillus terreus Involved in Plant
618 Biomass Degradation. Frontiers in Bioengineering and Biotechnology 8,
619 https://doi.org/10.3389/fbioe.2020.564527.

620 Courtine, D., Provaznik, J., Reboul, J., Blanc, G., Benes, V., Ewbank, J.J., 2020. Long-read only
621 assembly of Drechmeria coniospora genomes reveals widespread chromosome plasticity and
622 illustrates the limitations of current nanopore methods. GigaScience 9,
623 https://doi.org/10.1093/gigascience/giaa099.

624 Couturier, M., Ladevèze, S., Sulzenbacher, G., Ciano, L., Fanuel, M., Moreau, C., Villares, A.,
625 Cathala, B., Chaspoul, F., Frandsen, K.E., Labourel, A., Herpoël-Gimbert, I., Grisel, S., Haon,
626 M., Lenfant, N., Rogniaux, H., Ropartz, D., Davies, G.J., Rosso, M.-N., Walton, P.H.,
627 Henrissat, B., Berrin, J.-G., 2018. Lytic xylan oxidases from wood-decay fungi unlock
628 biomass degradation. Nature Chemical Biology 14, 306-310.
629 https://doi.org/10.1038/nchembio.2558.

630 Crucello, A., Sforça, D.A., Horta, M.A.C., dos Santos, C.A., Viana, A.J.C., Beloti, L.L., de Toledo,
631 M.A.S., Vincentz, M., Kuroshu, R.M., de Souza, A.P., 2015. Analysis of Genomic Regions of
632 Trichoderma harzianum IOC-3844 Related to Biomass Degradation. PLOS ONE 10,
633 e0122122. https://doi.org/10.1371/journal.pone.0122122.

634 Dal Molin, A., Minio, A., Griggio, F., Delledonne, M., Infantino, A., Aragona, M., 2018. The
635 genome assembly of the fungal pathogen Pyrenochaeta lycopersici from Single-Molecule
636 Real-Time sequencing sheds new light on its biological complexity. PLOS ONE 13,
637 e0200217. https://doi.org/10.1371/journal.pone.0200217.

638 Dana, C.M., Dotson-Fagerstrom, A., Roche, C.M., Kal, S.M., Chokhawala, H.A., Blanch, H.W.,
639 Clark, D.S., 2014. The importance of pyroglutamate in cellulase Cel7A. Biotechnology and
640 Bioengineering 111, 842-847. https://doi.org/https://doi.org/10.1002/bit.25178.

641 de Vries, R.P., Riley, R., Wiebenga, A., Aguilar-Osorio, G., Amillis, S., Uchima, C.A., Anderluh, G.,
642 Asadollahi, M., Askin, M., Barry, K., Battaglia, E., Bayram, Ö., Benocci, T., Braus-
643 Stromeyer, S.A., Caldana, C., Cánovas, D., Cerqueira, G.C., Chen, F., Chen, W., Choi, C.,
644 Clum, A., dos Santos, R.A.C., Damásio, A.R.d.L., Diallinas, G., Emri, T., Fekete, E., Flipphi,
645 M., Freyberg, S., Gallo, A., Gournas, C., Habgood, R., Hainaut, M., Harispe, M.L., Henrissat,
646 B., Hildén, K.S., Hope, R., Hossain, A., Karabika, E., Karaffa, L., Karányi, Z., Kraševec, N.,
647 Kuo, A., Kusch, H., LaButti, K., Lagendijk, E.L., Lapidus, A., Levasseur, A., Lindquist, E.,
648 Lipzen, A., Logrieco, A.F., MacCabe, A., Mäkelä, M.R., Malavazi, I., Melin, P., Meyer, V.,
649 Mielnichuk, N., Miskei, M., Molnár, Á.P., Mulé, G., Ngan, C.Y., Orejas, M., Orosz, E.,
650 Ouedraogo, J.P., Overkamp, K.M., Park, H.-S., Perrone, G., Piumi, F., Punt, P.J., Ram, A.F.J.,
651 Ramón, A., Rauscher, S., Record, E., Riaño-Pachón, D.M., Robert, V., Röhrig, J., Ruller, R.,
652 Salamov, A., Salih, N.S., Samson, R.A., Sándor, E., Sanguinetti, M., Schütze, T., Sepčić, K.,
653 Shelest, E., Sherlock, G., Sophianopoulou, V., Squina, F.M., Sun, H., Susca, A., Todd, R.B.,
654 Tsang, A., Unkles, S.E., van de Wiele, N., van Rossen-Uffink, D., Oliveira, J.V.d.C., Vesth,
655 T.C., Visser, J., Yu, J.-H., Zhou, M., Andersen, M.R., Archer, D.B., Baker, S.E., Benoit, I.,
656 Brakhage, A.A., Braus, G.H., Fischer, R., Frisvad, J.C., Goldman, G.H., Houbraken, J.,
657 Oakley, B., Pócsi, I., Scazzocchio, C., Seiboth, B., vanKuyk, P.A., Wortman, J., Dyer, P.S.,
658 Grigoriev, I.V., 2017. Comparative genomics reveals high biological diversity and specific
659 adaptations in the industrially and medically important fungal genus Aspergillus. Genome
660 Biology 18, 28. https://doi.org/10.1186/s13059-017-1151-0.

661 Delabona, P.D.S., Codima, C.A., Ramoni, J., Zubieta, M.P., de Araujo, B.M., Farinas, C.S., Pradella,
662 J., Seiboth, B., 2020a. The impact of putative methyltransferase overexpression on the
663 *Trichoderma harzianum* cellulolytic system for biomass conversion. Bioresour Technol 313,
664 123616. https://doi.org/10.1016/j.biortech.2020.123616.

665 Delabona, P.D.S., Codima, C.A., Ramoni, J., Zubieta, M.P., de Araujo, B.M., Farinas, C.S., Pradella,
666 J., Seiboth, B., 2020b. The impact of putative methyltransferase overexpression on the
667 *Trichoderma harzianum* cellulolytic system for biomass conversion. Bioresour. Technol. 313,
668 123616. https://doi.org/10.1016/j.biortech.2020.123616.

669 Dodd, S.L., Lieckfeldt, E., Samuels, G.J., 2003. Hypocrea atroviridis sp. nov., the teleomorph of
670 Trichoderma atroviride. Mycologia 95, 27-40.
671 https://doi.org/10.1080/15572536.2004.11833129.

672 Druzhinina, I.S., Chenthamara, K., Zhang, J., Atanasova, L., Yang, D., Miao, Y., Rahimi, M.J.,
673 Grujic, M., Cai, F., Pourmehdi, S., Salim, K.A., Pretzer, C., Kopchinskiy, A.G., Henrissat, B.,
674 Kuo, A., Hundley, H., Wang, M., Aerts, A., Salamov, A., Lipzen, A., LaButti, K., Barry, K.,
675 Grigoriev, I.V., Shen, Q., Kubicek, C.P., 2018a. Massive lateral transfer of genes encoding
676 plant cell wall-degrading enzymes to the mycoparasitic fungus Trichoderma from its plant-
677 associated hosts. PLoS genetics 14, e1007322-e1007322.
678 https://doi.org/10.1371/journal.pgen.1007322.

679 Druzhinina, I.S., Chenthamara, K., Zhang, J., Atanasova, L., Yang, D., Miao, Y., Rahimi, M.J.,
680 Grujic, M., Cai, F., Pourmehdi, S., Salim, K.A., Pretzer, C., Kopchinskiy, A.G., Henrissat, B.,

Kuo, A., Hundley, H., Wang, M., Aerts, A., Salamov, A., Lipzen, A., LaButti, K., Barry, K., Grigoriev, I.V., Shen, Q., Kubicek, C.P., 2018b. Massive lateral transfer of genes encoding plant cell wall-degrading enzymes to the mycoparasitic fungus Trichoderma from its plant-associated hosts. PLOS Genetics 14, e1007322. https://doi.org/10.1371/journal.pgen.1007322.

Druzhinina, I.S., Kopchinskiy, A.G., Kubicek, C.P., 2006. The first 100 Trichoderma species characterized by molecular data. Mycoscience 47, 55-64. https://doi.org/10.1007/s10267-006-0279-7.

Druzhinina, I.S., Kubicek, C.P., Komoń-Zelazowska, M., Mulaw, T.B., Bissett, J., 2010. The Trichoderma harzianum demon: complex speciation history resulting in coexistence of hypothetical biological species, recent agamospecies and numerous relict lineages. BMC Evolutionary Biology 10, 94. https://doi.org/10.1186/1471-2148-10-94.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology 16, 157. https://doi.org/10.1186/s13059-015-0721-2.

Emms, D.M., Kelly, S., 2017. STRIDE: Species Tree Root Inference from Gene Duplication Events. Molecular Biology and Evolution 34, 3267-3278. https://doi.org/10.1093/molbev/msx259.

Emms, D.M., Kelly, S., 2018. STAG: Species Tree Inference from All Genes. bioRxiv 267914. https://doi.org/10.1101/267914.

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology 20, 238. https://doi.org/10.1186/s13059-019-1832-y.

Fanelli, F., Liuzzi, V.C., Logrieco, A.F., Altomare, C., 2018. Genomic characterization of Trichoderma atrobrunneum (T. harzianum species complex) ITEM 908: insight into the genetic endowment of a multi-target biocontrol strain. BMC Genomics 19, 662. https://doi.org/10.1186/s12864-018-5049-3.

Fang, H., Li, C., Zhao, J., Zhao, C., 2021. Biotechnological Advances and Trends in Engineering Trichoderma reesei towards Cellulase Hyperproducer. Biotechnology and Bioprocess Engineering 26, 517-528. https://doi.org/10.1007/s12257-020-0243-y.

Ferreira Filho, J.A., Horta, M.A.C., Beloti, L.L., dos Santos, C.A., de Souza, A.P., 2017. Carbohydrate-active enzymes in Trichoderma harzianum: a bioinformatic analysis bioprospecting for key enzymes for the biofuels industry. BMC Genomics 18, 779. https://doi.org/10.1186/s12864-017-4181-9.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic acids research 39, W29-W37. https://doi.org/10.1093/nar/gkr367.

Fraceto, L.F., Maruyama, C.R., Guilger, M., Mishra, S., Keswani, C., Singh, H.B., de Lima, R., 2018. Trichoderma harzianum-based novel formulations: potential applications for management of Next-Gen agricultural challenges. Journal of Chemical Technology & Biotechnology 93, 2056-2063. https://doi.org/https://doi.org/10.1002/jctb.5613.

718 Gan, X., Cao, D., Zhang, Z., Cheng, S., Wei, L., Li, S., Liu, B., 2020. Draft Genome Assembly of
719 Floccularia luteovirens, an Edible and Symbiotic Mushroom on Qinghai-Tibet Plateau. G3:
720 Genes|Genomes|Genetics 10, 1167. https://doi.org/10.1534/g3.120.401037.

721 Ganesh Kumar, A., Manisha, D., Sujitha, K., Magesh Peter, D., Kirubagaran, R., Dharani, G., 2021.
722 Genome sequence analysis of deep sea Aspergillus sydowii BOBA1 and effect of high
723 pressure on biodegradation of spent engine oil. Scientific Reports 11, 9347.
724 https://doi.org/10.1038/s41598-021-88525-9.

725 Guo, B., Sato, N., Biely, P., Amano, Y., Nozaki, K., 2016. Comparison of catalytic properties of
726 multiple β-glucosidases of Trichoderma reesei. Applied Microbiology and Biotechnology
727 100, 4959-4968. https://doi.org/10.1007/s00253-016-7342-x.

728 Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for
729 genome assemblies. Bioinformatics (Oxford, England) 29, 1072-1075.
730 https://doi.org/10.1093/bioinformatics/btt086.

731 Hagestad, O.C., Hou, L., Andersen, J.H., Hansen, E.H., Altermark, B., Li, C., Kuhnert, E., Cox, R.J.,
732 Crous, P.W., Spatafora, J.W., Lail, K., Amirebrahimi, M., Lipzen, A., Pangilinan, J.,
733 Andreopoulos, W., Hayes, R.D., Ng, V., Grigoriev, I.V., Jackson, S.A., Sutton, T.D.S.,
734 Dobson, A.D.W., Rämä, T., 2021. Genomic characterization of three marine fungi, including
735 Emericellopsis atlantica sp. nov. with signatures of a generalist lifestyle and marine biomass
736 degradation. IMA Fungus 12, 21. https://doi.org/10.1186/s43008-021-00072-0.

737 Han, L., Tan, Y., Ma, W., Niu, K., Hou, S., Guo, W., Liu, Y., Fang, X., 2020. Precision Engineering
738 of the Transcription Factor Cre1 in Hypocrea jecorina (Trichoderma reesei) for Efficient
739 Cellulase Production in the Presence of Glucose. Frontiers in bioengineering and
740 biotechnology 8, 852-852. https://doi.org/10.3389/fbioe.2020.00852.

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

745 Horta, M.A.C., Thieme, N., Gao, Y., Burnum-Johnson, K.E., Nicora, C.D., Gritsenko, M.A., Lipton,
746 M.S., Mohanraj, K., de Assis, L.J., Lin, L., Tian, C., Braus, G.H., Borkovich, K.A., Schmoll,
747 M., Larrondo, L.F., Samal, A., Goldman, G.H., Benz, J.P., 2019. Broad Substrate-Specific
748 Phosphorylation Events Are Associated With the Initial Stage of Plant Cell Wall Recognition
749 in Neurospora crassa. Frontiers in Microbiology 10,
750 https://doi.org/10.3389/fmicb.2019.02317.

751 Horta, M.A.C., Vicentini, R., Delabona, P.d.S., Laborda, P., Crucello, A., Freitas, S., Kuroshu, R.M.,
752 Polikarpov, I., Pradella, J.G.d.C., Souza, A.P., 2014. Transcriptome Profile of Trichoderma
753 harzianum IOC-3844 Induced by Sugarcane Bagasse. PLOS ONE 9, e88689.
754 https://doi.org/10.1371/journal.pone.0088689.

755 Institute, B., Picard Tools. Picard. Broad Institute, GitHub repository.

Kappel, L., Münsterkötter, M., Sipos, G., Escobar Rodriguez, C., Gruber, S., 2020. Chitin and chitosan remodeling defines vegetative development and Trichoderma biocontrol. PLOS Pathogens 16, e1008320. https://doi.org/10.1371/journal.ppat.1008320.

Kidwai, M.K., Nehra, M., 2017. Biotechnological applications of *Trichoderma species* for environmental and food security, in: Gahlawat, S.K., Salar, R.K., Siwach, P., Duhan, J.S., Kumar, S., Kaur, P. (Eds.), Plant biotechnology: recent advancements and developments. Springer, Singapore, pp. 125-156.

Kjærbølling, I., Vesth, T., Frisvad, J.C., Nybo, J.L., Theobald, S., Kildgaard, S., Petersen, T.I., Kuo, A., Sato, A., Lyhne, E.K., Kogle, M.E., Wiebenga, A., Kun, R.S., Lubbers, R.J.M., Mäkelä, M.R., Barry, K., Chovatia, M., Clum, A., Daum, C., Haridas, S., He, G., LaButti, K., Lipzen, A., Mondo, S., Pangilinan, J., Riley, R., Salamov, A., Simmons, B.A., Magnuson, J.K., Henrissat, B., Mortensen, U.H., Larsen, T.O., de Vries, R.P., Grigoriev, I.V., Machida, M., Baker, S.E., Andersen, M.R., 2020. A comparative genomics study of 23 Aspergillus species from section Flavi. Nature Communications 11, 1106. https://doi.org/10.1038/s41467-019-14051-y.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome research 27, 722-736. https://doi.org/10.1101/gr.215087.116.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. Journal of Molecular Biology 305, 567-580. https://doi.org/https://doi.org/10.1006/jmbi.2000.4315.

Kubicek, C.P., Herrera-Estrella, A., Seidl-Seiboth, V., Martinez, D.A., Druzhinina, I.S., Thon, M., Zeilinger, S., Casas-Flores, S., Horwitz, B.A., Mukherjee, P.K., Mukherjee, M., Kredics, L., Alcaraz, L.D., Aerts, A., Antal, Z., Atanasova, L., Cervantes-Badillo, M.G., Challacombe, J., Chertkov, O., McCluskey, K., Coulpier, F., Deshpande, N., von Döhren, H., Ebbole, D.J., Esquivel-Naranjo, E.U., Fekete, E., Flipphi, M., Glaser, F., Gómez-Rodríguez, E.Y., Gruber, S., Han, C., Henrissat, B., Hermosa, R., Hernández-Oñate, M., Karaffa, L., Kosti, I., Le Crom, S., Lindquist, E., Lucas, S., Lübeck, M., Lübeck, P.S., Margeot, A., Metz, B., Misra, M., Nevalainen, H., Omann, M., Packer, N., Perrone, G., Uresti-Rivera, E.E., Salamov, A., Schmoll, M., Seiboth, B., Shapiro, H., Sukno, S., Tamayo-Ramos, J.A., Tisch, D., Wiest, A., Wilkinson, H.H., Zhang, M., Coutinho, P.M., Kenerley, C.M., Monte, E., Baker, S.E., Grigoriev, I.V., 2011. Comparative genome sequence analysis underscores mycoparasitism as the ancestral life style of Trichoderma. Genome Biology 12, R40. https://doi.org/10.1186/gb-2011-12-4-r40.

Kubicek, C.P., Steindorff, A.S., Chenthamara, K., Manganiello, G., Henrissat, B., Zhang, J., Cai, F., Kopchinskiy, A.G., Kubicek, E.M., Kuo, A., Baroncelli, R., Sarrocco, S., Noronha, E.F., Vannacci, G., Shen, Q., Grigoriev, I.V., Druzhinina, I.S., 2019. Evolution and comparative genomics of the most common Trichoderma species. BMC Genomics 20, 485. https://doi.org/10.1186/s12864-019-5680-7.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biology 5, R12. https://doi.org/10.1186/gb-2004-5-2-r12.

799 Letunic, I., Bork, P., 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree
800 display and annotation. Bioinformatics 23, 127-128.
801 https://doi.org/10.1093/bioinformatics/btl529.

802 Levasseur, A., Drula, E., Lombard, V., Coutinho, P.M., Henrissat, B., 2013. Expansion of the
803 enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes.
804 Biotechnology for Biofuels 6, 41. https://doi.org/10.1186/1754-6834-6-41.

805 Li, B., Walton, J.D., 2017. Functional diversity for biomass deconstruction in family 5 subfamily 5
806 (GH5_5) of fungal endo-β1,4-glucanases. Applied Microbiology and Biotechnology 101,
807 4093-4101. https://doi.org/10.1007/s00253-017-8168-x.

808 Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics (Oxford,
809 England) 34, 3094-3100. https://doi.org/10.1093/bioinformatics/bty191.

810 Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform.
811 Bioinformatics (Oxford, England) 26, 589-595.
812 https://doi.org/10.1093/bioinformatics/btp698.

813 Li, W.-C., Huang, C.-H., Chen, C.-L., Chuang, Y.-C., Tung, S.-Y., Wang, T.-F., 2017. Trichoderma
814 reesei complete genome sequence, repeat-induced point mutation, and partitioning of
815 CAZyme gene clusters. Biotechnology for Biofuels 10, 170. https://doi.org/10.1186/s13068-
816 017-0825-x.

817 Mahmoud, M., Gobet, N., Cruz-Dávalos, D.I., Mounier, N., Dessimoz, C., Sedlazeck, F.J., 2019.
818 Structural variant calling: the long and the short of it. Genome Biology 20, 246.
819 https://doi.org/10.1186/s13059-019-1828-7.

820 Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., Zimin, A., 2018. MUMmer4:
821 A fast and versatile genome alignment system. PLOS Computational Biology 14, e1005944.
822 https://doi.org/10.1371/journal.pcbi.1005944.

823 Martinez, D., Berka, R.M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S.E., Chapman, J.,
824 Chertkov, O., Coutinho, P.M., Cullen, D., Danchin, E.G.J., Grigoriev, I.V., Harris, P.,
825 Jackson, M., Kubicek, C.P., Han, C.S., Ho, I., Larrondo, L.F., de Leon, A.L., Magnuson, J.K.,
826 Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A.A., Schmoll, M.,
827 Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C.L., Yao, J., Barabote, R., Nelson,
828 M.A., Detter, C., Bruce, D., Kuske, C.R., Xie, G., Richardson, P., Rokhsar, D.S., Lucas, S.M.,
829 Rubin, E.M., Dunn-Coleman, N., Ward, M., Brettin, T.S., 2008. Genome sequencing and
830 analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).
831 Nature Biotechnology 26, 553-560. https://doi.org/10.1038/nbt1403.

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

836 Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C.,
837 Ye, K., Cheetham, R.K., Chinwalla, A., Conrad, D.F., Fu, Y., Grubert, F., Hajirasouliha, I.,
838 Hormozdiari, F., Iakoucheva, L.M., Iqbal, Z., Kang, S., Kidd, J.M., Konkel, M.K., Korn, J.,

Khurana, E., Kural, D., Lam, H.Y.K., Leng, J., Li, R., Li, Y., Lin, C.-Y., Luo, R., Mu, X.J., Nemesh, J., Peckham, H.E., Rausch, T., Scally, A., Shi, X., Stromberg, M.P., Stütz, A.M., Urban, A.E., Walker, J.A., Wu, J., Zhang, Y., Zhang, Z.D., Batzer, M.A., Ding, L., Marth, G.T., McVean, G., Sebat, J., Snyder, M., Wang, J., Ye, K., Eichler, E.E., Gerstein, M.B., Hurles, M.E., Lee, C., McCarroll, S.A., Korbel, J.O., Genomes, P., 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470, 59-65. https://doi.org/10.1038/nature09708.

Mitsuhashi, S., Matsumoto, N., 2020. Long-read sequencing for rare human genetic diseases. Journal of Human Genetics 65, 11-19. https://doi.org/10.1038/s10038-019-0671-8.

Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sánchez-García, M., Morin, E., Andreopoulos, B., Barry, K.W., Bonito, G., Buée, M., Carver, A., Chen, C., Cichocki, N., Clum, A., Culley, D., Crous, P.W., Fauchery, L., Girlanda, M., Hayes, R.D., Kéri, Z., LaButti, K., Lipzen, A., Lombard, V., Magnuson, J., Maillard, F., Murat, C., Nolan, M., Ohm, R.A., Pangilinan, J., Pereira, M.d.F., Perotto, S., Peter, M., Pfister, S., Riley, R., Sitrit, Y., Stielow, J.B., Szöllősi, G., Žifčáková, L., Štursová, M., Spatafora, J.W., Tedersoo, L., Vaario, L.-M., Yamada, A., Yan, M., Wang, P., Xu, J., Bruns, T., Baldrian, P., Vilgalys, R., Dunand, C., Henrissat, B., Grigoriev, I.V., Hibbett, D., Nagy, L.G., Martin, F.M., 2020. Large-scale genome sequencing of mycorrhizal fungi provides insights into the early evolution of symbiotic traits. Nature Communications 11, 5125. https://doi.org/10.1038/s41467-020-18795-w.

Monclaro, A.V., Filho, E.X.F., 2017. Fungal lytic polysaccharide monooxygenases from family AA9: Recent developments and application in lignocelullose breakdown. International Journal of Biological Macromolecules 102, 771-778. https://doi.org/https://doi.org/10.1016/j.ijbiomac.2017.04.077.

Montoliu-Nerin, M., Sánchez-García, M., Bergin, C., Grabherr, M., Ellis, B., Kutschera, V.E., Kierczak, M., Johannesson, H., Rosling, A., 2020. Building de novo reference genome assemblies of complex eukaryotic microorganisms from single nuclei. Scientific Reports 10, 1303. https://doi.org/10.1038/s41598-020-58025-3.

Motta, M.L.L., Filho, J.A.F., de Melo, R.R., Zanphorlin, L.M., dos Santos, C.A., de Souza, A.P., 2021. A novel fungal metal-dependent α-l-arabinofuranosidase of family 54 glycoside hydrolase shows expanded substrate specificity. Scientific Reports 11, 10961. https://doi.org/10.1038/s41598-021-90490-2.

Nagel, J.H., Wingfield, M.J., Slippers, B., 2021. Increased abundance of secreted hydrolytic enzymes and secondary metabolite gene clusters define the genomes of latent plant pathogens in the Botryosphaeriaceae. BMC Genomics 22, 589. https://doi.org/10.1186/s12864-021-07902-w.

Najjarzadeh, N., Matsakas, L., Rova, U., Christakopoulos, P., 2021. How Carbon Source and Degree of Oligosaccharide Polymerization Affect Production of Cellulase-Degrading Enzymes by Fusarium oxysporum f. sp. lycopersici. Frontiers in Microbiology 12, https://doi.org/10.3389/fmicb.2021.652655.

Nakkeeran, S., Marimuthu, T., Renukadevi, P., Brindhadevi, S., Jogaiah, S., 2021. 24 - Exploring the biogeographical diversity of Trichoderma for plant health, in: Jogaiah, S. (Ed.), Biocontrol Agents and Secondary Metabolites. Woodhead Publishing, pp. 537-571.

Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics (Oxford, England) 32, 292-294. https://doi.org/10.1093/bioinformatics/btv566.

Priest, S.J., Yadav, V., Heitman, J., 2020. Advances in understanding the evolution of fungal genome architecture. F1000Research 9, F1000 Faculty Rev-1776. https://doi.org/10.12688/f1000research.25424.1.

Ramazi, S., Zahiri, J., 2021. Post-translational modifications in proteins: resources, tools and prediction methods. Database 2021, baab012. https://doi.org/10.1093/database/baab012.

Rosolen, R.R., Aono, A.H., Almeida, D.A., Filho, J.A.F., Horta, M.A.C., de Souza, A.P., 2021. Network analysis reveals different strategies of Trichoderma spp. associated with XYR1 and CRE1 during cellulose degradation. bioRxiv 2020.2005.2002.074344. https://doi.org/10.1101/2020.05.02.074344.

Sakamoto, Y., Sereewattanawoot, S., Suzuki, A., 2020. A new era of long-read sequencing for cancer genomics. Journal of Human Genetics 65, 3-10. https://doi.org/10.1038/s10038-019-0658-5.

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.

Schmoll, M., 2018. Regulation of plant cell wall degradation by light in Trichoderma. Fungal Biology and Biotechnology 5, 10. https://doi.org/10.1186/s40694-018-0052-7.

Schmoll, M., Dattenböck, C., Carreras-Villaseñor, N., Mendoza-Mendoza, A., Tisch, D., Alemán, M.I., Baker, S.E., Brown, C., Cervantes-Badillo, M.G., Cetz-Chel, J., Cristobal-Mondragon, G.R., Delaye, L., Esquivel-Naranjo, E.U., Frischmann, A., Gallardo-Negrete, J.d.J., García-Esquivel, M., Gomez-Rodriguez, E.Y., Greenwood, D.R., Hernández-Oñate, M., Kruszewska, J.S., Lawry, R., Mora-Montes, H.M., Muñoz-Centeno, T., Nieto-Jacobo, M.F., Nogueira Lopez, G., Olmedo-Monfil, V., Osorio-Concepcion, M., Piłsyk, S., Pomraning, K.R., Rodriguez-Iglesias, A., Rosales-Saavedra, M.T., Sánchez-Arreguín, J.A., Seidl-Seiboth, V., Stewart, A., Uresti-Rivera, E.E., Wang, C.-L., Wang, T.-F., Zeilinger, S., Casas-Flores, S., Herrera-Estrella, A., 2016. The Genomes of Three Uneven Siblings: Footprints of the Lifestyles of Three Trichoderma Species. Microbiology and molecular biology reviews : MMBR 80, 205-327. https://doi.org/10.1128/MMBR.00040-15.

Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., Schatz, M.C., 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nature methods 15, 461-468. https://doi.org/10.1038/s41592-018-0001-7.

Sharma, S., Kour, D., Rana, K.L., Dhiman, A., Thakur, S., Thakur, P., Thakur, S., Thakur, N., Sudheer, S., Yadav, N., Yadav, A.N., Rastegari, A.A., Singh, K., 2019. Trichoderma: Biodiversity, Ecological Significances, and Industrial Applications, in: Yadav, A.N., Mishra, S., Singh, S., Gupta, A. (Eds.), Recent Advancement in White Biotechnology Through Fungi: Volume 1: Diversity and Enzymes Perspectives. Springer International Publishing, Cham, pp. 85-120.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212. https://doi.org/10.1093/bioinformatics/btv351.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., Morgenstern, B., 2006. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Research 34, W435-W439. https://doi.org/10.1093/nar/gkl200.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., Koonin, E.V., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic acids research 28, 33-36. https://doi.org/10.1093/nar/28.1.33.

Thanh, V.N., Thuy, N.T., Huong, H.T.T., Hien, D.D., Hang, D.T.M., Anh, D.T.K., Hüttner, S., Larsbrink, J., Olsson, L., 2019. Surveying of acid-tolerant thermophilic lignocellulolytic fungi in Vietnam reveals surprisingly high genetic diversity. Scientific reports 9, 3674-3674. https://doi.org/10.1038/s41598-019-40213-5.

The UniProt, C., 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research 49, D480-D489. https://doi.org/10.1093/nar/gkaa1100.

Törönen, P., Medlar, A., Holm, L., 2018. PANNZER2: a rapid functional annotation web server. Nucleic acids research 46, W84-W88. https://doi.org/10.1093/nar/gky350.

van Eerde, A., Várnai, A., Jameson, J.K., Paruch, L., Moen, A., Anonsen, J.H., Chylenski, P., Steen, H.S., Heldal, I., Bock, R., Eijsink, V.G.H., Liu-Clarke, J., 2020. In-depth characterization of Trichoderma reesei cellobiohydrolase TrCel7A produced in Nicotiana benthamiana reveals limitations of cellulase production in plants by host-specific post-translational modifications. Plant biotechnology journal 18, 631-643. https://doi.org/10.1111/pbi.13227.

Varga, T., Krizsán, K., Földi, C., Dima, B., Sánchez-García, M., Sánchez-Ramírez, S., Szöllősi, G.J., Szarkándi, J.G., Papp, V., Albert, L., Andreopoulos, W., Angelini, C., Antonín, V., Barry, K.W., Bougher, N.L., Buchanan, P., Buyck, B., Bense, V., Catcheside, P., Chovatia, M., Cooper, J., Dämon, W., Desjardin, D., Finy, P., Geml, J., Haridas, S., Hughes, K., Justo, A., Karasiński, D., Kautmanova, I., Kiss, B., Kocsubé, S., Kotiranta, H., LaButti, K.M., Lechner, B.E., Liimatainen, K., Lipzen, A., Lukács, Z., Mihaltcheva, S., Morgado, L.N., Niskanen, T., Noordeloos, M.E., Ohm, R.A., Ortiz-Santana, B., Ovrebo, C., Rácz, N., Riley, R., Savchenko, A., Shiryaev, A., Soop, K., Spirin, V., Szebenyi, C., Tomšovský, M., Tulloss, R.E., Uehling, J., Grigoriev, I.V., Vágvölgyi, C., Papp, T., Martin, F.M., Miettinen, O., Hibbett, D.S., Nagy, L.G., 2019. Megaphylogeny resolves global patterns of mushroom evolution. Nature ecology & evolution 3, 668-678. https://doi.org/10.1038/s41559-019-0834-1.

Wei, H., Wu, M., Fan, A., Su, H., 2021. Recombinant protein production in the filamentous fungus Trichoderma. Chinese Journal of Chemical Engineering 30, 74-81. https://doi.org/https://doi.org/10.1016/j.cjche.2020.11.006.

Wu, J.Q., Dong, C., Song, L., Park, R.F., 2020. Long-Read–Based de novo Genome Assembly and Comparative Genomics of the Wheat Leaf Rust Pathogen Puccinia triticina Identifies Candidates for Three Avirulence Genes. Frontiers in Genetics 11, https://doi.org/10.3389/fgene.2020.00521.

Wu, L., McCluskey, K., Desmeth, P., Liu, S., Hideaki, S., Yin, Y., Moriya, O., Itoh, T., Kim, C.Y., Lee, J.-S., Zhou, Y., Kawasaki, H., Hazbón, M.H., Robert, V., Boekhout, T., Lima, N., Evtushenko, L., Boundy-Mills, K., Bunk, B., Moore, E.R.B., Eurwilaichitr, L., Ingsriswang, S., Shah, H., Yao, S., Jin, T., Huang, J., Shi, W., Sun, Q., Fan, G., Li, W., Li, X., Kurtböke, İ., Ma, J., 2018. The global catalogue of microorganisms 10K type strain sequencing project: closing the genomic gaps for the validly published prokaryotic and fungi species. GigaScience 7, https://doi.org/10.1093/gigascience/giy026.

Xie, B.-B., Qin, Q.-L., Shi, M., Chen, L.-L., Shu, Y.-L., Luo, Y., Wang, X.-W., Rong, J.-C., Gong, Z.-T., Li, D., Sun, C.-Y., Liu, G.-M., Dong, X.-W., Pang, X.-H., Huang, F., Liu, W., Chen, X.-L., Zhou, B.-C., Zhang, Y.-Z., Song, X.-Y., 2014. Comparative Genomics Provide Insights into Evolution of Trichoderma Nutrition Style. Genome Biology and Evolution 6, 379-390. https://doi.org/10.1093/gbe/evu018.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., Yin, Y., 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 46, W95-W101. https://doi.org/10.1093/nar/gky418.

Zhang, Y., Yang, J., Luo, L., Wang, E., Wang, R., Liu, L., Liu, J., Yuan, H., 2020a. 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.

Zhang, Y., Yang, J., Luo, L., Wang, E., Wang, R., Liu, L., Liu, J., Yuan, H., 2020b. 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.

Zin, N.A., Badaluddin, N.A., 2020. Biological functions of Trichoderma spp. for agriculture applications. Annals of Agricultural Sciences 65, 168-178. https://doi.org/https://doi.org/10.1016/j.aoas.2020.09.003.

Figure legends

Figure 1. *Trichoderma* isolates evaluated in this study. (**A**) Th3844, (**B**) Th0179, (**C**) Tr0711, and (**D**) Ta0020 cultivated on potato dextrose broth (PDA) solid medium at 28 °C. After DNA extraction, their genome was sequenced and assembled. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844; Tr0711: *T. reesei* CBMAI-0711.

Figure 2. Comparisons between the genomes of the analyzed *Trichoderma* isolates and the *T. reesei* QM6a reference genome. Dot plots of the assemblies of (**A**) Th3844, (**B**) Th0179, (**C**) Ta0020, and (**D**) Tr0711 that were generated by Canu (y-axis) against those from *T. reesei* QM6a (x-axis) that are available in the NCBI database. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844; Tr0711: *T. reesei* CBMAI-0711.

Figure 3. COG functional category distribution of the *Trichoderma* isolates considered. The plot shows the number of genes in the genomes of (**A**) Th3844, (**B**) Th0179, (**C**) Ta0020, and (**D**) Tr0711, in which a COG classification was obtained. The size of the boxes represents the abundance of the genes at the level of individual COG families. Only the COG functional categories with more than a hundred counts were represented. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844; Tr0711: *T. reesei* CBMAI-0711.

Figure 4. Distribution of CAZymes in *Trichoderma* spp. (**A**) The predicted CAZymes from the assembled genomes were classified according to the CAZy database. (**B**) The secreted CAZymes were grouped according to their CAZyme class. CAZymes: carbohydrate-active enzymes; Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844; Tr0711: *T. reesei* CBMAI-0711. AA: auxiliary activity; CBM: carbohydrate-binding module; EC: carbohydrate esterase; GH: glycosyl hydrolase; GT: glycosyl transferase; PL: polysaccharide lyase.

Figure 5. Quantitative comparison of the biomass-degrading enzyme repertoires of *Trichoderma* isolates. Heatmap of the number of enzymes in each CAZY family from the Th3844, Th0179, Ta0020, and Tr0711 genomes. This map includes only the enzymes/proteins related to biomass degradation. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844; Tr0711: *T. reesei* CBMAI-0711; LPMOs: Lytic polysaccharide monooxygenases; CDHs: Cellobiose dehydrogenases; PHDs: Pyranose dehydrogenases.

Figure 6. Phylogenetic relationships of *Trichoderma* spp. as inferred by an orthology analysis. The phylogenetic tree modeled by OrthoFinder software was based on the concatenation of 2,229 single copy orthogroups. In addition to the proteomes of Th3844, Th0179, Ta0020, and Tr0711, this methodology shows the inferred relationships among 19 *Trichoderma* spp., for which the proteomes are available in the NCBI database. *Fusarium* spp., *Aspergillus* spp., and *Neurospora* spp. were used as the outgroup. Bootstrap values are shown at the nodes.

Figure 7. Structural variant heterogeneity across the genomes of the evaluated *Trichoderma* spp. (**A**) Long-read alignment-based structural variant (SV) analyses among the evaluated *Trichoderma* isolates and *T. reesei* QM6a showed breakends (BNDs), deletions (DELs), multiple nucleotides and InDels (MIXEDs), duplications (DUPs), insertions (INSs), and inversions (INVs) between the genomes. (**B**) Functional effects of the identified SVs. Ta0020: *T. atroviride* CBMAI-0020; Th0179: *T. harzianum* CBMAI-0179; Th3844: *T. harzianum* IOC-3844; Tr0711: *T. reesei* CBMAI-0711.

Tables

Table 1. Genome assembly and annotation statistics.

	Th3844	**Th0179**	**Ta0020**	**Tr0711**
Number of reads	418,031	504,913	458,142	282,252
Genome size (bp)	40,219,724	39,170,259	36,411,897	32,448,670
GC content (%)	47.5	49.4	49.5	53.5
N50 reads	3,607,994	2,983,622	3,146,023	1,694,659
L50 contigs	5	6	5	7
Number of contigs	15	18	14	26
Complete BUSCOs (%)	90.1	98.7	99	99.1
Complete and single-copy BUSCOs (%)	89.6	98.1	98.8	98.9
Complete and duplicated BUSCOs (%)	0.5	0.6	0.2	0.2
Fragmented BUSCOs (%)	0.2	0.2	0.1	0.1
Missing (%)	9.7	1.1	0.9	0.8
Number of predicted genes	10,786	11,322	10,082	8,796
Number of annotated genes	10,611	11,065	9,547	8,495

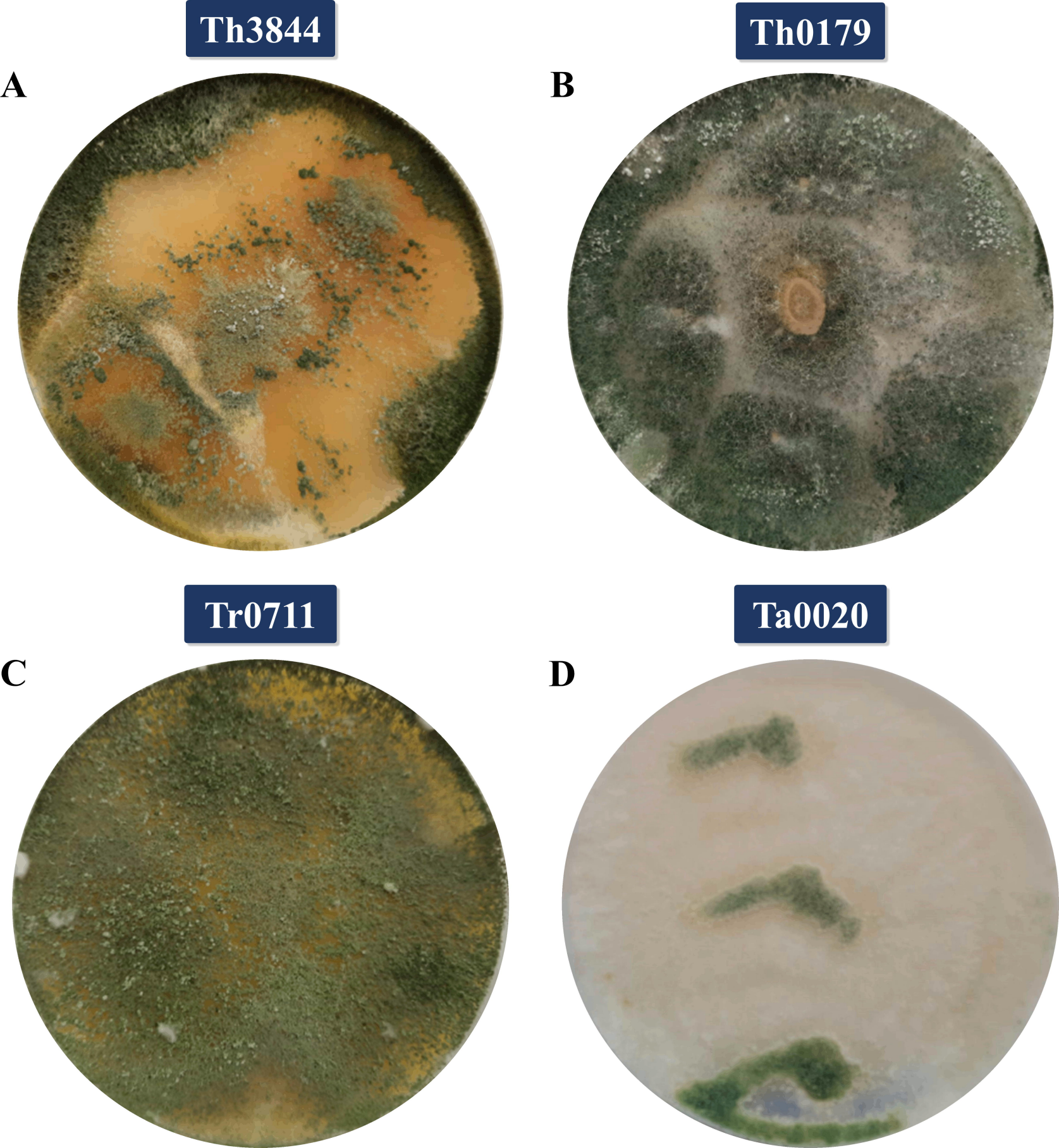
Table 2. Comparison of the genome features of *Trichoderma* spp. genomes.

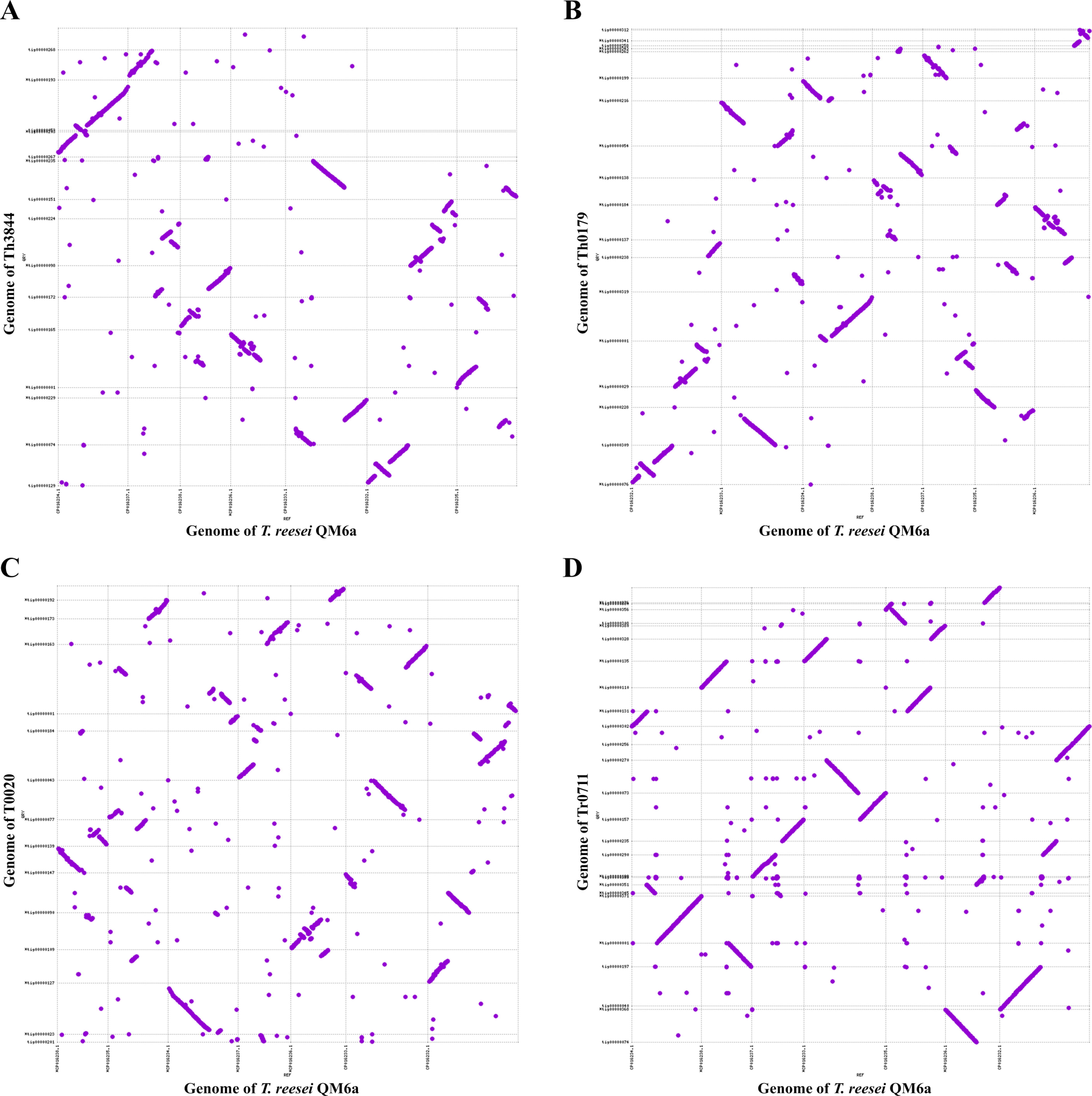
Species	Strain	Coverage	NCBI accession	Genome size (Mb)	GC content (%)	N50 reads	L50 contigs	Assembly level	Genes	Reference
T. harzianum	IOC-3844	164 x	SAMN23309297	40	47.5	3,607,994	5	Contigs (15)	10,786	This study
T. harzianum	CBMAI-0179	219 x	SAMN23309298	39	49.4	2,983,622	6	Contigs (18)	11,322	This study
T. harzianum	T6776	85 x	SAMN02851310	39	48.5	68,846	178	Scaffolds (1,572)	11,501	(Baroncelli et al., 2015)
T. harzianum	CBS 226.95	120 x	SAMN00761861	41	47.6	2,414,909	7	Scaffolds (532)	14,269	-
T. virens	Gv29–8	8 x	SAMN02744059	39	49.2	1,836,662	8	Scaffolds (93)	12,405	(Kubicek et al., 2011)
T. simmonsii	GH-Sj1	-	SAMN15516371	40	48.13	6,451,197	3	Scaffolds (7)	13,296	(Chung et al., 2021)
T. atroviride	CBMAI-0020	229 x	SAMN23309299	36	49.5	3,146,023	5	Contigs (14)	10,082	This study
T. atroviride	IMI206040	8 x	SAMN02744066	36	49.7	2,007,903	6	Contigs (29)	11,809	(Kubicek et al., 2011)
T. reesei	CBMAI-0711	143 x	SAMN23309300	32	53.5	1,694,659	7	Contigs (26)	8,796	This study
T. reesei	QM6a	80 x	SAMN05250858	35	51.0	18,236	3	Chromosomes (7)	10,877	(Li et al., 2017)
T. reesei	QM6a	9 x	SAMN02746107	33	52.8	1,219,543	9	Scaffolds (77)	9,109	(Martinez et al., 2008)

1074 **Table 3.** Distribution of Th3844, Th0179, Ta0020, and Tr0711 orthologs.

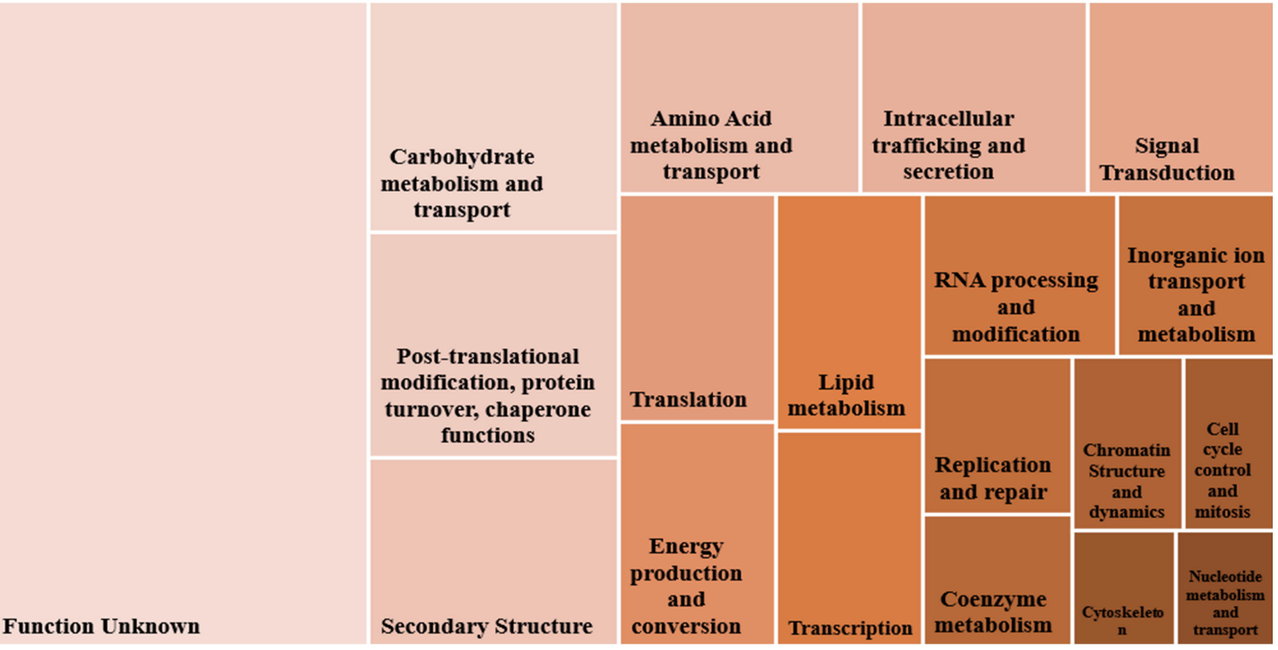
	Th3844	**Th0179**	**Ta0020**	**Tr0711**
Th3844	-	9,729	7,927	7,330
Th0179	9,729	-	8,755	8,141
Ta0020	7,927	8,755	-	7,784
Tr0711	7,330	8,141	7,784	-

1075

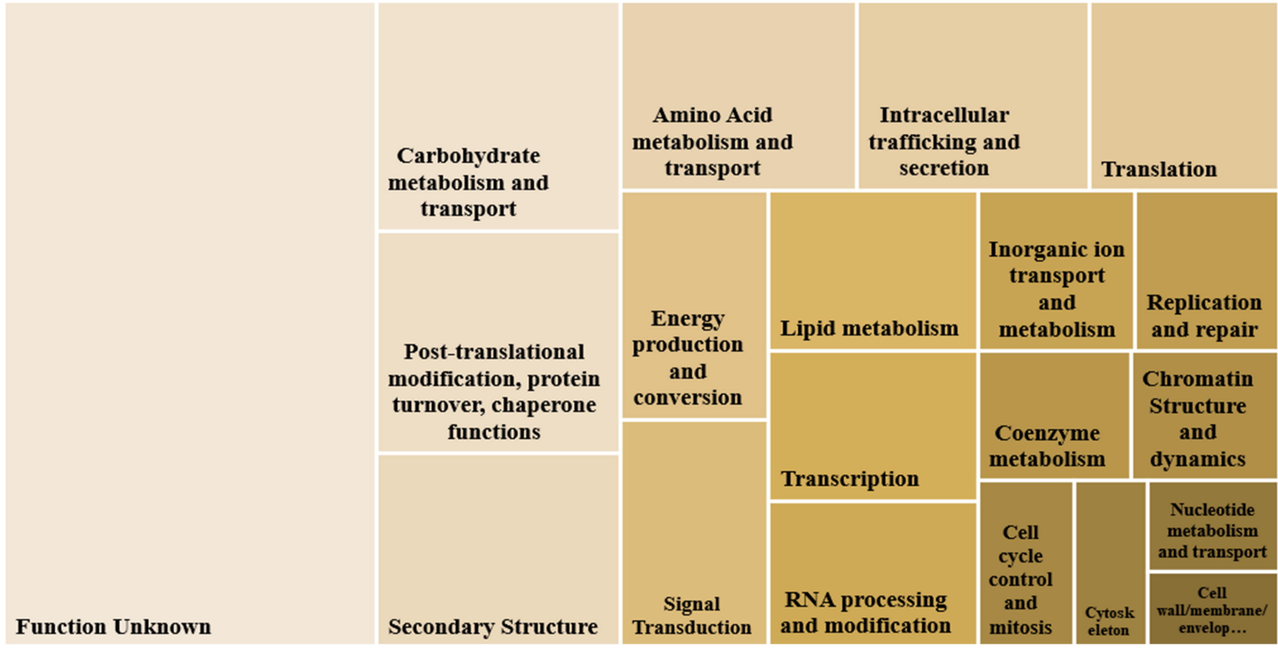

Th3844
A
Th0179
B
Tr0711
C
Ta0020
D


A
Genome of Th3844
Genome of T. reesei QM6a
B
Genome of Th0179
Genome of T. reesei QM6a
C
Genome of T0020
Genome of T. reesei QM6a
D
Genome of Tr0711
Genome of T. reesei QM6a

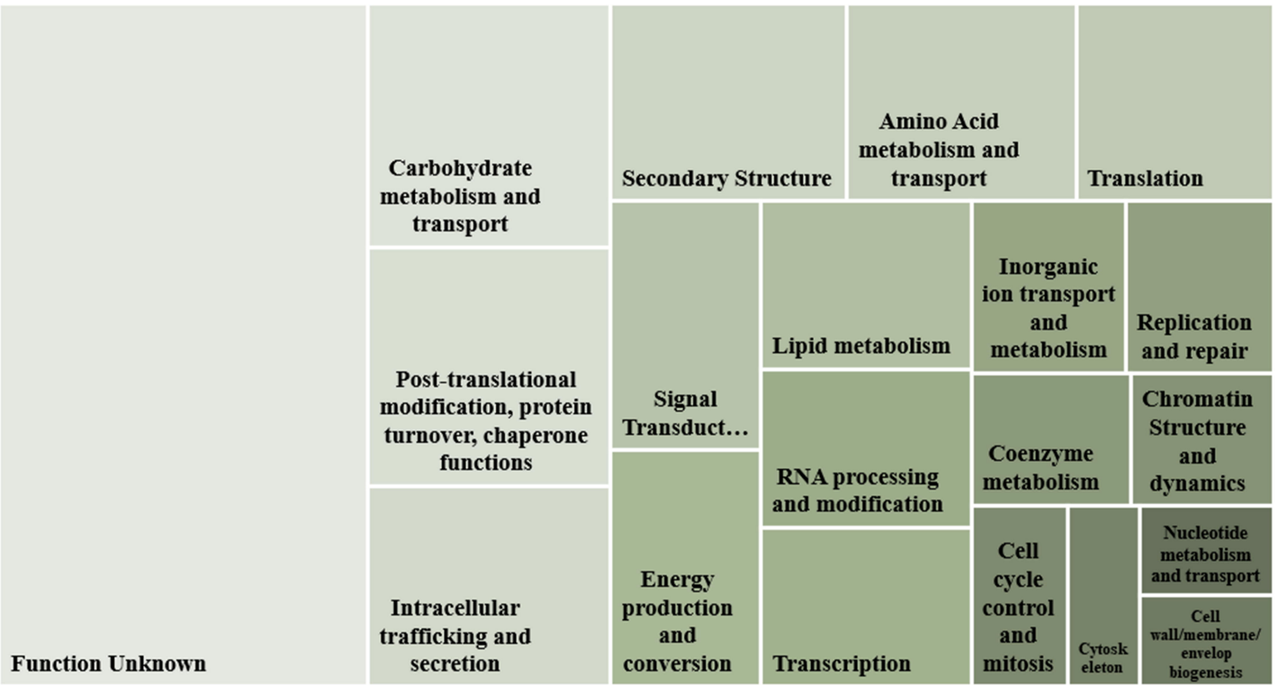
A


Th3844
Function Unknown
Carbohydrate metabolism and transport
Post-translational modification, protein turnover, chaperone functions
Secondary Structure
Amino Acid metabolism and transport
Intracellular trafficking and secretion
Signal Transduction
Translation
Energy production and conversion
Lipid metabolism
Transcription
RNA processing and modification
Inorganic ion transport and metabolism
Replication and repair
Coenzyme metabolism
Chromatin Structure and dynamics
Cytoskeleton
Cell cycle control and mitosis
Nucleotide metabolism and transport

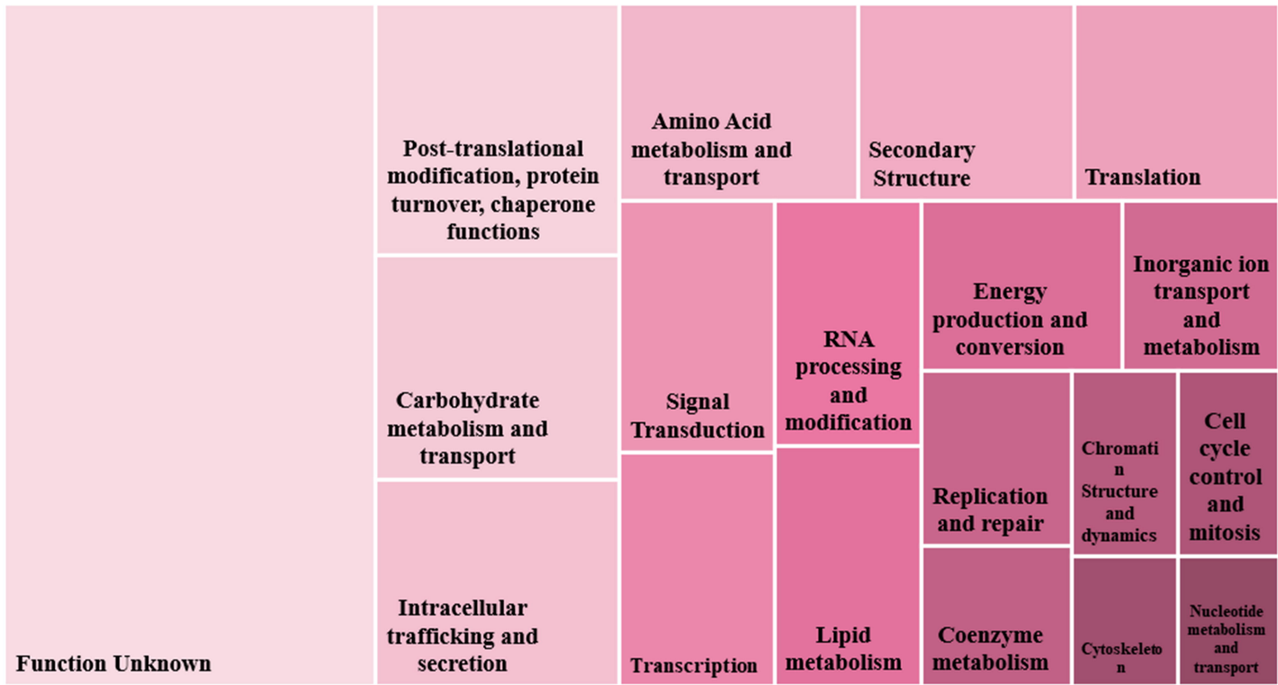
B

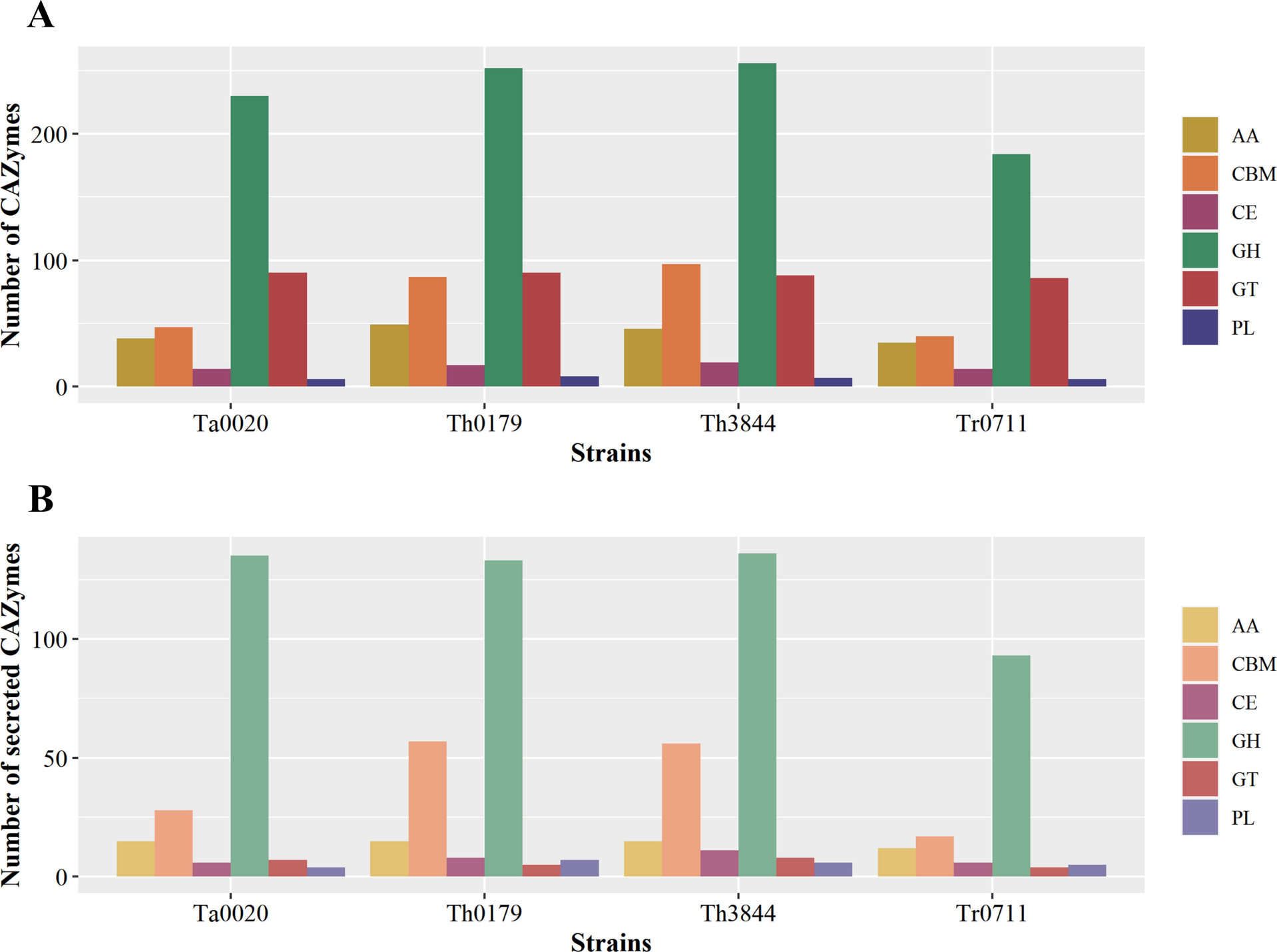

Th0179
Function Unknown
Carbohydrate metabolism and transport
Post-translational modification, protein turnover, chaperone functions
Secondary Structure
Amino Acid metabolism and transport
Intracellular trafficking and secretion
Translation
Energy production and conversion
Signal Transduction
Lipid metabolism
Transcription
RNA processing and modification
Inorganic ion transport and metabolism
Coenzyme metabolism
Replication and repair
Chromatin Structure and dynamics
Cell cycle control and mitosis
Cytoskeleton
Nucleotide metabolism and transport
Cell wall/membrane/ envelop…

C

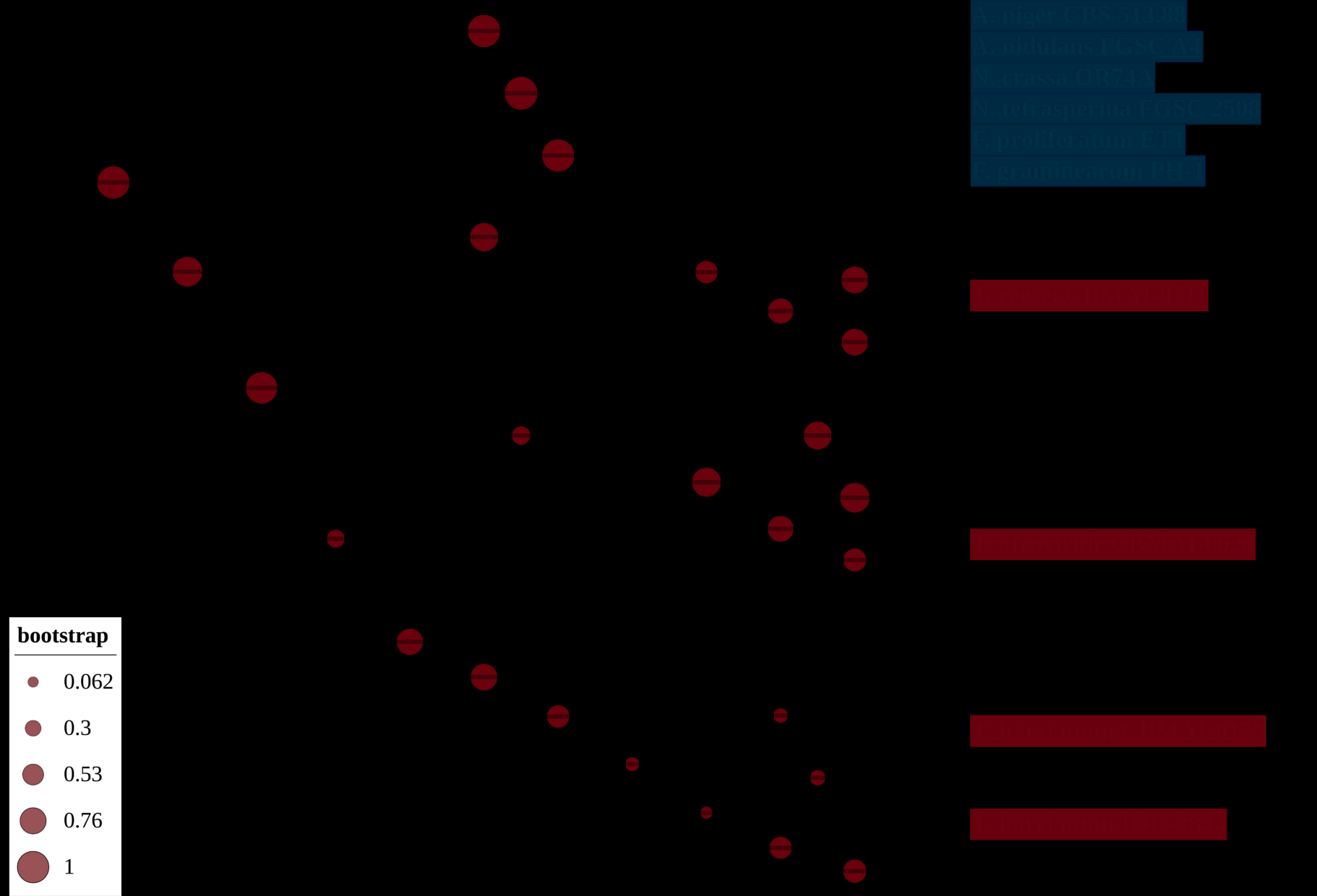

Ta0020
Function Unknown
Carbohydrate metabolism and transport
Post-translational modification, protein turnover, chaperone functions
Intracellular trafficking and secretion
Secondary Structure
Amino Acid metabolism and transport
Translation
Signal Transduct…
Energy production and conversion
Lipid metabolism
Transcription
RNA processing and modification
Inorganic ion transport and metabolism
Coenzyme metabolism
Replication and repair
Chromatin Structure and dynamics
Cell cycle control and mitosis
Cytoskeleton
Nucleotide metabolism and transport
Cell wall/membrane/ envelop biogenesis

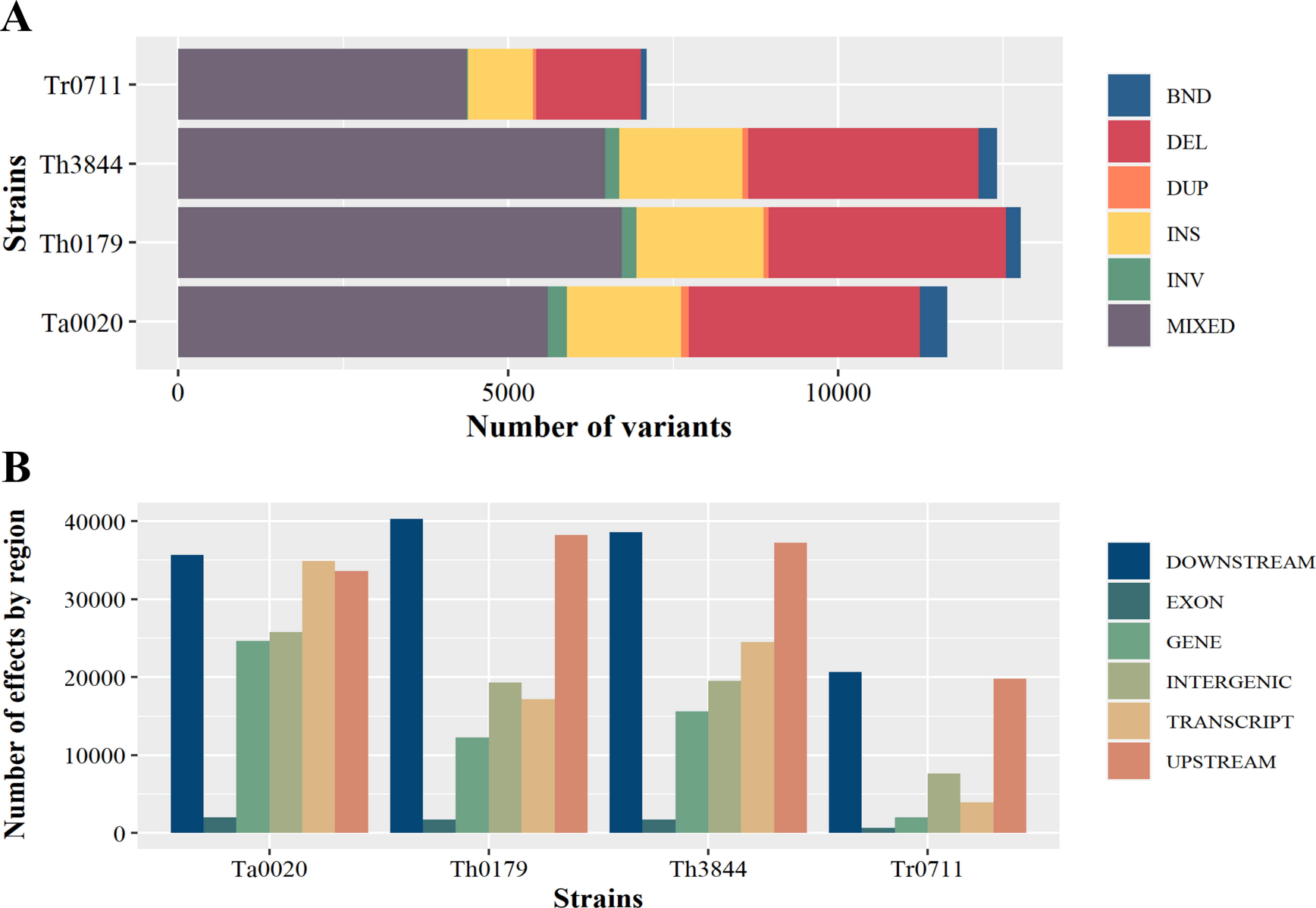
D


Tr0711
Function Unknown
Post-translational modification, protein turnover, chaperone functions
Carbohydrate metabolism and transport
Intracellular trafficking and secretion
Amino Acid metabolism and transport
Secondary Structure
Translation
Signal Transduction
Transcription
RNA processing and modification
Lipid metabolism
Energy production and conversion
Inorganic ion transport and metabolism
Replication and repair
Coenzyme metabolism
Chromatin Structure and dynamics
Cell cycle control and mitosis
Cytoskeleton
Nucleotide metabolism and transport


A
Number of CAZymes
200
100
0
Ta0020
Th0179
Th3844
Tr0711
Strains
AA
CBM
CE
GH
GT
PL
B
Number of secreted CAZymes
100
50
0
Ta0020
Th0179
Th3844
Tr0711
Strains
AA
CBM
CE
GH
GT
PL

Substrate/enzyme category	Enzyme class	CAZy family	Th3844	Th0179	Ta0020	Tr0711
Arabinan	Exo-arabinanases/Arabinofuranosidases	GH51	0	0	1	0
	Arabinofuranosidases	GH54	2	2	2	2
	Endoarabinanases	GH43	5	4	4	2
Cellulose	Cellobiohydrolases	GH6	1	1	1	1
	Oligosaccharide oxidases	AA7	2	3	1	1
	LPMOs	AA9	3	3	3	3
	Endoglucanases	GH5	10	10	9	7
Esterases	Feruloyl/p-coumaroyl/acetyl esterases	CE1	1	0	0	0
	4-O-methyl-glucoronoyl esterases	CE15	1	1	1	1
	Acetyl esterases	CE5	4	4	3	4
Lignin	Glyoxal oxidases (GLOX)	AA5	1	1	1	1
	1,4-benzoquinone reductases	AA6	1	1	1	1
	Vanillyl-alcohol oxidases	AA4	2	3	0	0
	Peroxidases	AA2	1	3	5	3
	Laccases	AA1	9	9	6	4
Lignin/Cellulose	Oxidoreductases/CDHs	AA3	19	20	12	12
Monosugars	PDHs	AA12	1	1	1	0
Pectin	Pectin lyases	PL1	0	0	1	0
	Pectin methylesterases	CE8	1	1	1	0
	Rhamnosidases	GH78	2	2	2	1
	Pectin lyases	PL20	2	2	2	2
	Polygalacturonases	GH28	5	5	5	4
Soluble oligosaccharides	β-glucosidases	GH1	4	4	4	2
	Mannosidases	GH2	11	13	10	7
Xylan	Xylanases	GH10	2	2	1	1
	LPMOs	AA14	2	2	2	2
	Xylanases	GH11	4	4	4	3
Xylosidases/β-glucosidases	Xylosidases/β-glucosidases	GH3	16	17	15	12


A. niger CBS 513.88
A. nidulans FGSC A4
N. crassa OR74A
N. tetrasperma FGSC 2508
F. proliferatum ET1
F. graminearum PH-1
bootstrap
0.062
0.3
0.53
0.76
1


A
Tr0711
Th3844
Th0179
Ta0020
Strains
0
5000
10000
Number of variants
BND
DEL
DUP
INS
INV
MIXED
B
Number of effects by region
40000
30000
20000
10000
0
Ta0020
Th0179
Th3844
Tr0711
Strains
DOWNSTREAM
EXON
GENE
INTERGENIC
TRANSCRIPT
UPSTREAM