Comparative genomics uncovers genetic diversity and synthetic biology of secondary metabolite production of the genus Trametes

CURRENT STATUS: POSTED

Yan Zhang
Hefei Normal University

zhangyanwind@163.com
Corresponding Author
ORCiD: https://orcid.org/0000-0001-9914-1302

Jie Cao
Nanjing Biotogether Co., Ltd

Jingjing Wang
Hefei Normal University

Yajun Cheng
Hefei Normal University

Minghui Zhou
Hefei Normal University

Wei Wang
Hefei Normal University

Ming Geng
Hefei Normal University

Decong Xu
Hefei Normal University

Zhongdong Xu
Hefei Normal University

DOI:
10.21203/rs.2.11205/v1

SUBJECT AREAS
Epigenetics & Genomics  Molecular Biology
Abstract
It’s well-established that the CAZyme genes of genus Trametes contributed to the degradation processes of polysaccharides, including lignin or crystalline cellulose. However, the comprehensive analysis of the composition of CAZymes and the biosynthetic gene clusters of Trametes genus remain unclear. We conducted comparative analysis, detected the CAZyme genes, and predicted the biosynthetic gene clusters for 9 Trametes strains. Among 82,053 homologous clusters we obtained for genus Trametes, we identified 8,518 core genes, 60,441 accessory genes and 13,094 specific genes. Our results showed that a large proportion of CAZyme genes were catalogued into glycoside hydrolases, glycosyltransferases, and carbohydrate esterases. The predicted BGCs of Trametes genus were divided into 6 strategies and the 9 Trametes strains harbored 47.78 BGCs on average. Our study uncovers the genus Trametes exhibited an open pan-genome structure, provides insights into the genetic diversity and explores the synthetic biology of secondary metabolite production for Trametes genus.

Introduction
The genus *Trametes*, one of white-rot basidiomycetes, belonging to the family Polyporaceae and the class Agaricomycetes of fungi commonly grows in the tiled layer of decaying wood \[1\] and uses deciduous tree \[2\]. *Trametes* genus currently comprises at least 20 species until April 2018 \[3, 4\], such as *Trametes* sp., *Trametes trogii* and *Trametes villosa* CCMB561. The genus *Trametes* is known to be decolorize dyes \[1\] and several species of genus *Trametes* have been used for centuries in traditional medicine in East Asia counties \[5\]. A previous study has reported that *Trametes* sp. SQ01 has the ability to decolorize azo, anthraquinone, and triphenylmethane dyes \[6\]. *T. trogii*, which is a worldwide-distributed white-rot fungi and an outstanding laccase producer, degrade all lignocellulosic materials \[6\].
7. *T. villosa* CCMB561 was isolated from the decaying wood and a common species in the Brazilian semiarid region [8]. It is well-established that *T. villosa* is a high potential fungus for biotechnological applications due to the fact that it is a good producer of the three important ligninolytic enzymes (laccase, manganese peroxidase and lignin peroxidase) [4]. Moreover, the laccases from the *Trametes* genus have been reported as a biocatalyst in enzymatic biofuel cells [9, 10]. Hence, understanding the genetic diversity of *Trametes* genus is beneficial for its biotechnological applications.

Carbohydrate-active enzymes (known as CAZymes) are involved in the degradation processes of polysaccharides of plant cell wall [11], lignin, cellulose, hemicellulose and pectin [12]. Especially, the CAZyme classes (glycoside hydrolases, carbohydrate eaterases, and polysaccharide lyases) play central roles in plant biomass decomposition by bacteria and fungi [13]. It has been reported that the total expression levels of CAZyme genes of *T. versicolor* F21a were almost twice as that of the control group during the algicidal process and the majority of expressed CAZyme genes belonged to glycoside hydrolases family and auxiliary activities [14]. Similarly, 590 genes of *T. villosa* CCMB561 were identified as CAZyme genes [4]. Although the compositions of CAZyme genes of fewer species of the *Trametes* genus have been reported, the comprehensive analysis of the composition of CAZyme genes of *Trametes* genus still remains unclear.

Secondary metabolites (referred to as natural products or specialized metabolites) are the foundation of many drugs [15] and are also important chemicals widely used in agriculture and nutrition [16]. It is well-known that fungi are rich sources of thousands of second metabolites (SMs) and fungal SMs can be grouped into four main chemical types: polyketides (PKS), terpenoids, shikimic
acid derived compounds, and non-ribosomal peptides (NRPS) [17]. Secondary metabolites of fungal are crucial players in its development and actively shape interactions with other microbes [18]. A previous study has reported that the number of genes coded for secondary metabolism of ascomycetes is higher than that of basidiomycetes, archeo-ascomycetes, and chytridiomycetes [19]. In recent years, lot of fungal genomes were sequenced and the development of various genome mining software tools, such as antiSMASH [16] and ClusterFinder [20], enabled researchers to analyze the biosynthetic gene clusters (BGCs) of secondary metabolites. It has been reported that the genomes of filamentous fungi contain up to 90 potential BGCs encoding their diverse secondary metabolites [21] and 24 genomes of *Penicillium* genus were mined for BGCs and associated PKS and NRPS BGCs to known pathways [22]. However, there has been no systemically comparative analysis on the identification of BGCs of secondary metabolite in *Trametes* genus.

In this study, we collected 9 available genomic data from 8 species of *Trametes* genus and conducted a comparative genomics analysis to investigate the genetic diversity and the synthetic biology of secondary metabolite. We obtained the pan-genome of *Trametes* genus, annotated the pan-genome against COG database and eggNOG database, and detected the CAZyme genes. The biosynthetic gene clusters for 9 *Trametes* strains were also predicted. Our results showed that the genus *Trametes* exhibited an open pan-genome structure and diverse genetic diversity for *Trametes* strains. The different distribution of CAZymes of *Trametes* genus revealed that the difference of ability for utilizing the carbohydrates for *Trametes* strains. The predicted biosynthetic gene clusters with unknown functions in *Trametes* genus suggested that *Trametes* genus has a great potential value for producing secondary metabolite.

Results And Discussions
Pan-genome construction and analysis

To conduct the pan-genome analysis, 9 available genomic data from 8 species within the genus Trametes were collected. The assembled genome of Trametes genus ranged from 31.62 Mb to 57.98 Mb, while the number of contigs/scaffolds ranged from 13 to 10,327, as well as the number of CDSs/ORFs predicting by Pordigal ranged from 56,735 to 101,817 (Table 1). To characterize the differences of genomic composition among these 9 strains, we detected the orthologs from the 610,421 high-quality proteins of Trametes genus. In total, we obtained 82,053 homologous clusters and the accumulation curve of homologous cluster showed that Trametes genus exhibited an open pan-genome structure (Figure 1a). The size of pan-genome of Trametes genus tended to increase gradually with new strains and was estimated to be 82,053 non-redundant genes within 9 strains (Figure 1a). While the size of core-genome tended to decrease progressively with new strains and was comprised of 8,518 non-redundant genes within 9 strains of Trametes genus (Figure 1a).

In contrast, we obtained 60,441 accessory genes and 13,094 specific genes of Trametes genus. The distribution of accessory genes varied from 17,587 to 31,856. Although the size of T. cinnabarina FP 104138-Sp is smallest (Table 1) among 9 strains of Trametes genus, it had the maximum number (31,856) of accessory genes. T. hirsute had the minimum number (17,587) of accessory genes. The number of specific orthologs of 9 Trametes strains ranged from 368 to 4,698, in which T. villoosa CCMB561 clearly stood out possessing the highest number (4,698) of specific genes, whereas T. cinnabarina FP 104138-Sp had the minimum number (368) of specific genes (Figure 1b). The distribution of accessory genes and specific genes revealed that the genomes of Trametes genus are considerably diverse.

Phylogenetic analysis of Trametes pan-genome

It is well-established that the nuclear ribosomal Internal Transcribed Spacer (ITS) region is the primary fungal barcode marker [23] and useful to identify the broadest range of fungi [24]. To analyze the phylogenetic relationships for 9 Trametes strains, we built the phylogenetic tree for Trametes genus and Coriolopsis caperata, which was selected as outgroup to
root the topology of the tree, based on their ITS1 and ITS2 sequences, respectively. We observed that the ITS1 and ITS2 sequences have a strong power to differentiate the 9 strains of Trametes genus (Figure 1c-1d). Among the phylogenetic trees, we found that T. pubescens, T. versicolor FP-101664 SS1 and T. sp. AH28-2 have closer evolutionary relationship, and T. cinnabarina BRFM137 has closer evolutionary relationship with T. coccinea BRFM310. However, the evolutionary relationship of T. hirsuta was more distant from that of remaining strains of Trametes genus (Figure 1c-1d).

To gain insights into the evolutionary relationships among the strains of Trametes genus, we conducted the phylogenetic tree based on the genes presence and absence in pan-genome. The result showed that T. villosa has closer evolutionary relationship with and T. versicolor FP-101664 SS1, while T. sp. AH28-2, T. pubescens, T. coccinea BRFM310, and T. cinnabarina FP 104138-Sp have closer evolutionary relationship with each other (Figure 1e), which revealed that these strains have similar genetic composition and suggested that these strains might have similar functional metabolism. Meanwhile, the result also revealed that the evolutionary relationship of T. hirsuta was distant from that of rest of strains of Trametes genus (Figure 1e), which suggested that the functional composition of T. hirsuta is different from remaining strains of Trametes genus and the functional metabolism of Trametes genus is more diverse.

**Functional characterization of the Trametes pan-genome**

To investigate the functional diversity of 9 strains of Trametes genus, we annotated the pan-genome of Trametes genus against the Cluster of Orthologous Group (COG) to characterize functions of the accessory clusters, core clusters and specific genes. It was obvious that the orthologous genes of the pan-genome were mainly dominant in “general function prediction only” (R, 1,758, 15.47%), “replication, recombination and repair” (L, 998, 8.78%), “amino acid transport and metabolism” (E, 996, 8.77%), “translation, ribosomal structure and biogenesis” (J, 755, 6.64%), and “carbohydrate transport and metabolism” (G, 753, 6.63%, Figure 2a). Moreover, we observed that Trametes genus has 442 (3.89%) orthologous genes assigned to “secondary metabolites biosynthesis, transport and catabolism” (Q, Figure 2a). Similarly, the accessory genome was enriched in “general function prediction only” (R, 1,411, 15.29%), “replication, recombination and repair” (L, 810, 8.78%) and
“amino acid transport and metabolism” (E, 799, 8.66%, Figure 2a). In addition, the core gene clusters were predominantly composed of genes involved to “general function prediction only” (R, 88, 20.09%), “energy production and conversion” (C, 51, 11.64%), “amino acid transport and metabolism” (E, 38, 8.68%), and “carbohydrate transport and metabolism” (G, 36, 8.21%, Figure 2a). Finally, the specific genes were mainly contributed to the functions involved in “replication, recombination and repair” (L, 610, 28.44%) and “general function prediction only” (R, 259, 12.07%, Figure 2a). We speculated that the high proportions of accessory genes (9,230, 81.23%) and specific genes (2,145, 18.88%) in each strain of Trametes genus might lead to a more diverse of functions. To explore the evolution of function catalogues, we also conducted a GO analysis to characterize genetic functions of the pan-genome. These genes were categorized according to biological process, cellular component and molecular function. The results showed that a great number of genes of pan-genome involved in various enzymes, metabolic pathways and biological processes. Specifically, the enrichment analysis of biological processes showed that many genes were classified into three categories: biosynthetic process (2,265), response to stress (2,186), and small molecule metabolic process (1,882, Figure 2b). In addition, a total of 606 orthologous genes were contributed to the carbohydrate metabolic process (Figure 2b). The classification according to molecular function showed that the majority of genes grouped into binding (such as ion binding, DNA binding, and RNA binding) and enzymes (such as ATPase activity, oxidoreductase activity, and kinase activity), yet various genes also enriched in other categories included transporter activity and enzyme regulator activity (Figure 2b). The enrichment analysis of cellular component showed that the higher proportions of genes were grouped into the functions, including cell (8,492), intracellular (8,386), and cytoplasm (7,363, Figure 2b). Among these categories, we observed that accessory clusters are highly represented in these functions than core and specific genes (Figure 2b). Thus, we speculated that the diversity of accessory clusters might influence their functional diversity, which allows these strains to utilize the resources of surrounding environment and better adapt to the environment.

Identification of CAZymes for Trametes genus
It is well-established that the white-rot basidiomycetes have ability to decompose lignin most
efficiently [25]. The genus *Trametes*, as one of branches of white-rot basidiomycetes, the members of this genus, such as *T. villosa* [25] and *T. gibbosa* BRFM 952 [26], have been reported that these strains have an unexpected high activity on lignin or crystalline cellulose. Previous study has showed that the carbohydrate-active enzymes (CAZymes) play a central role in the degradation process of glycoconjugate, oligo- and polysaccharide [27]. Hence, to obtain systematic understanding of the CAZymes of *Trametes* genus, we systemically identified the pan-genome of *Trametes* genus against the CAZy database and grouped CAZyme genes into different CAZyme families and CBMs based on family-specific HMMs [28]. In total, we identified 280 orthologous genes (Figure 3a) and grouped these genes into 87 CAZyme families. Our results showed that a large proportion of orthologous genes catalogued into glycoside hydrolases (35.36%, GHs), glycosyltransferases (21.07%, GTs), and carbohydrate esterases (13.57%, CEs, Figure 3a). The GH hydrolyze the glycosidic bond between two or more carbohydrates [27] and it contributes the most catalytic enzymes and involves in the degradation of lignocelluloses [29]. Moreover, we observed that the number of CAZymes of accessory clusters are higher than that of core clusters and specific genes (Figure 3a) and the number of genes belonged to different CAZymes in each *Trametes* strain (Figure 3b) are different, which suggested that the ability of strains of *Trametes* genus in the degradation of carbohydrates have significant difference.

To gain insights into the CAZyme composition of 9 strains of *Trametes* genus, we subdivided the 87 CAZyme families based on accessory clusters, core clusters and specific genes (Figure 3c-3e). As to the orthologous genes of accessory clusters, a slice of CAZyme families, such as AA9, CBM1, CE10, CE16, CE4, GH10, GH16, GH18, GH5, GH79, GT2, GT20, GT76 and PL14, were detected in all 9 *Trametes* strains (Figure 3c). Among these CAZyme families, we found that the number of CE10 (ranged from 5 to 20) and GH16 (ranged from 6 to 15) were higher than other CAZyme families (Figure 3c). We found that the counts of AA2 of *T. cinnabarina* BRFM137 (14) and CBM21 of *T. villosa*
CCMB561 (24) were higher than that of other strains (Figure 3c). In contrast, the distributions of CAZyme families of core clusters, including AA2, CBM1, CBM14, CBM50, GH18, GH3, GH37 and GT90, in Trametes strains were similar (Figure 3d). Additionally, the orthologous genes of specific genes were assigned to different CAZyme families (Figure 3e). For example, 14 CAZyme families, including GH18 and GT4, were dominant in T. villosa CCMB561, and CE10 and GH16 were enriched in T. sp. AH28-2. Interestingly, we observed that AA7 was only detected in T. villosa CCMB561, CE5 was only identified in T. versicolor FP-101664 SS1, CH6 only appeared in T. pubescens, and CH89 was only detected T. hirsute. In a summary, the composition of CAZyme families of 9 Trametes strains showed that these strains have the ability to metabolize the same carbohydrates and also have the potential ability to degrade different carbohydrates.

Identification of biosynthetic gene clusters (BGCs) of Trametes genus

To obtain a better understanding of the secondary metabolite of Trametes genus, we applied antiSMASH to predict the biosynthetic gene clusters (BGCs). Our results showed that the BGCs of Trametes genus were mainly divided into 6 strategies, including putative cluster of unknown type identified with the ClusterFinder algorithm (cf_putative), terpene cluster (terpene), nonribosomal peptide synthetase cluster (nrps), putative fatty acid cluster identified with the ClusterFinder algorithm (cf_fatty_acid), type I PKS cluster (t1pks), and lanthipeptide cluster (lantipeptide, Figure 4). Specifically, the 9 strains of Trametes genus harborod 47.78 BGCs on average and the number of BGCs of Trametes genus was ranged from 18 (T. cinnabarina FP 104138-Sp) to 66 (T. sp. AH28-2, Figure 4). Clusters of cf_putative and terpenes represented the majority of predicted specialized metabolite BGCs and contributed to the highest share among the 9 Trametes strains (Figure 4). The results showed that the clusters of cf_putative were enriched in the 9 strains of Trametes genus (ranged from 3 to 39, on average 25 BGCs, Figure 4), while T. cinnabarina FP 104138-Sp harbored only 3 clusters of cf_putative. When comparing the BGCs of terpene, it can be seen that T. hirsute harbored 17 BGCs and T. cinnabarina BRFM137 and T. cinnabarina FP 104138-Sp all harbored 9 BGCs. Moreover, a majority of these identified terpene BGCs could not be linked to known terpene BGCs. In addition, many predicted BGCs were grouped into “other” of these 9 strains.
Although the secondary metabolites of predicted BGCs are unknown to us, we identified quite a few enzymes as the core biosynthetic enzymes, such as cembrene C synthase, pristinol synthase, (+)-epi-cubebol synthase, and epi-isozizaene synthase, in several BGCs in most strains of Trametes genus. To gain insights of BGCs of Trametes genus, we chose three BGCs of T. pubescens as examples (Figure 5). Specifically, the BGCs of No. 13 was detected by the ClusterFinder algorithm built-in antiSMASH with 81.47% probability and the homologous gene clusters were identified from T. versicolor FP-101664 SS1 with 75% of genes show similarity and Dichomitus squalens LYAD-421 SS1 with 25% genes show similarity (Figure 5a). In addition, the monomer of predicted core structure of substrate was ala-ala (C₆H₁₂O₂N₂). The BGCs of No. 15 was detected as other unknown cluster with 83.63% probability and the homologous gene clusters were identified from T. cinnabarina BRFM137 with 66% of genes show similarity and T. versicolor FP-101664 SS1 with 66% of genes show similarity (Figure 5b), while the BGCs of No. 20 was also detected as other unknown cluster and the homologous gene clusters were identified from T. versicolor FP-101664 SS1 with 93% of genes show similarity and T. cinnabarina BRFM 137 with 56% of genes show similarity (Figure 5c). In summary, these results revealed that many BGCs of Trametes genus were undiscovered and the Trametes genus has a great potential value for producing secondary metabolite.

Conclusions
In conclusion, we systemically comparative analyzed the pan-genome of Trametes genus, one of important branches of white-rot basidiomycetes, based on the available genome of 9 Trametes strains and we explored the functional composition of 9 strains of Trametes genus, including COG annotation, GO annotation and CAZyme families. Furthermore, we predicted the biosynthetic gene clusters for each strain of Trametes genus. Our comparative analysis revealed that the Trametes genus exhibited an open pan-genome structure and tremendous diversity in both the number and variety of genes. The results of comparative analysis of pan-genome of Trametes genus also showed that the differences in the composition of functional metabolisms, as well as the evolutionary relationships among the 9 strains of Trametes genus. Additionally, we observed that the high proportion of accessory genes and specific genes annotated with different COG annotations and GO annotations
showed that the diversity of the functions among the *Trametes* strains of *Trametes* genus. Importantly, our results showed that the CAZyme families, such as GH, GT and CE, enriched in the *Trametes* strains, while the number of CAZymes of accessory clusters are higher than that of core clusters and specific genes and the number of genes belonged to different CAZymes in each *Trametes* strain suggested that the potential ability to degrade different carbohydrates for different strains of *Trametes* genus. Finally, we predicted many biosynthetic gene clusters with unknown functions in *Trametes* genus, which suggested that *Trametes* genus has a great potential value for producing secondary metabolite. Together, we uncover the genetic diversity and synthetic biology of secondary metabolite production of *Trametes* genus in this work. Our results provide a broader understanding of *Trametes* genus and importantly shed the lights on the composition of CAZyme families and the biosynthetic gene clusters of *Trametes* genus.

Materials And Methods
Genomic data collection of genus *Trametes*
We used ‘Trametes’ as the keyword and searched in genome database of National Center for Biotechnology Information (NCBI) on Sep 14th, 2018. We found that the genus *Trametes* comprising 8 species and containing 9 full genomes, such as *T. cinnabarina* BRFM137, *T. cinnabarina* FP 104138-Sp, *T. coccinea* BRFM310, *T. hirsuta*, *T. polyzona*, *T. pubescens*, *T. sp. AH28-2*, *T. versicolor* FP-101664 SS1 and *T. villosa*. Thus, we downloaded the genomic data (draft) of these strains for comparative genomics analysis.

Gene prediction and orthologs identification
We conducted the genome annotation for 9 strains using Pordigal [30] (version: 2.6). Specifically, to predict the genes and proteins for 9 genomic data of *Trametes* genus, we applied the Pordigal with default setting to recognize open reading frames (ORFs) and protein sequences [31]. We identified the protein orthologs of *Trametes* genus by using OrthMCL [32] (version: 2.0.9) with e-value < 1e-5 and inflation parameter of 1.5. We divided the homologous clusters into three groups: core, accessory and specific groups. The core genes or proteins represent the genes or proteins shared in all 9 genome of *Trametes* genus used in our study,
while the accessory genes or proteins comprised genes or proteins shared by at least two strains but not all 9 strains of *Trametes* genus. The rest of genes or proteins only occurred in one strain were clustered into specific groups (strain-specific genes or proteins). We obtained the protein sequences of homologous clusters for *Trametes* genus.

**Phylogeny analysis**

The nuclear ribosomal Internal Transcribed Spacer (ITS) region has been reported as the primary fungal barcode marker [23] and ITS1 or ITS2 was widely used to identify the broadest range of fungi [24]. Hence, we applied ITSx [33], which is an open source software utility to extract the highly variable ITS1 and ITS2 subregions from ITS sequences, to extract the ITS1 and ITS2 sequences from the genomic data of 9 strains of *Trametes* genus. Meanwhile, we chose the ITS1 and ITS2 sequences of *Coriolopsis caperata* (accession number: AB158316) as outgroup. In our work, due to the fact that the ITS1 and ITS2 sequences of *Trametes coccinea* were not extracted, we selected the ITS1 and ITS2 sequences of *Pycnoporus coccineus* strain MUCL 38523 [34] (one strain of *Trametes coccinea*, accession number: FJ873395) to represent *Trametes coccinea* and construct the phylogenetic tree. The ITS1 and ITS2 sequences were aligned using the MUSCLE (version: 3.8.31) [35] with default setting, respectively. The multiple sequences alignment result was eliminated by using Gblocks (version: 0.91b) [36] to remove the regions that were divergent, misaligned, or with a larger number of gaps. Then we used PHYLIP (version: 3.696) [37] with 100 bootstrap iterations to construct a maximum likelihood tree based on the concatenated multiple sequences alignment result. Moreover, we built a pan-genome tree by MEGA software (version: 5) [38] based on the composition of orthologous. Specifically, we calculated the pairwise Manhattan distance between each strain based on the presence and absence of orthologs proteins.
Functional annotation
To obtain the COG functions, the protein sequences of homologous clusters for *Trametes* genus were aligned to individual COG proteins [39], which downloaded from NCBI (https://www.ncbi.nlm.nih.gov/COG/) using BLSATP with e-value < 1e-4 and the top one of the result was chosen as the best annotation of each protein. The annotations of the protein sequences of homologous clusters were assigned to 25 functional categories. We also annotated the protein sequences of homologous clusters against eggNOG annotation database by using the eggnog-mapper [40] and we divided the GO annotations of protein sequences of homologous clusters into three groups, including biological process, molecular function and cellular component. We annotated these protein sequences against CAZyme database to investigate the functional composition related to carbohydrate metabolism. Specifically, we downloaded the CAZy database from dbCAN [41] (http://csbl.bmb.uga.edu/dbCAN/). According to the manual of dbCAN CAZyme annotation, protein sequences were annotated by running hmmscan from the HMMER (version: 3.1b1) [27, 42] with default setting and the annotated results was summarized into glycosyltransferase (GT), glycoside hydrolase (GH), carbohydrate-binding module (CBM), auxiliary activity (AA), polysaccharide lyase (PL) and carbohydrate esterase (CE).

Biosynthetic gene clusters detection
To detect the biosynthetic pathways of secondary metabolites for *Trametes* genus, antiSMASH (version: 4.0) [16] was used to identify the biosynthetic gene clusters (BGCs). A set of 9 genome sequences of *Trametes* genus was used as the input data for predicting the BGCs by using antiSMASH with the mode of fungi and other default setting. The BGCs were determined to their subtypes (e.g., type I polyketide, terpenoid).

Declarations

Authors’ contributions
This study was designed by YZ. YZ collected and downloaded the data. YZ and YJC analyzed the data. YZ, JJW, and ZDX wrote the initial draft of the manuscript. All authors revised the manuscript.

**Acknowledgments**

This work was partially supported by Natural Science Foundation of the Anhui Higher Education Institutions of China grant Kj2018A0494, Youth Elite Support Plan in Universities of Anhui Province grant gxyq2018056, National Science Foundation of China grant 31800049, Natural Science Foundation of the Anhui grant 1608085MC64.

**Additional Information**

Competing financial interests: The authors declare no competing financial interests.

**References**

1. Cho K-S, Ryu HW: Biodecolorization and biodegradation of dye by fungi: A review. *KSBB Journal* 2015, 30(5):203-222.

2. Ryu H, Ryu HW, Cho K-S: Characterization of Dye Decolorization in Cell-Free Culture Broth of Trametes versicolor CBR43. *Journal of microbiology and biotechnology* 2017, 27(1):155-160.

3. Justo A, Miettinen O, Floudas D, Ortiz-Santana B, Sjökvist E, Lindner D, Nakasone K, Niemelä T, Larsson K-H, Ryvarden L: A revised family-level classification of the Polyporales (Basidiomycota). *Fungal biology* 2017, 121(9):798-824.

4. Ferreira DSS, Kato RB, Miranda FM, da Costa Pinheiro K, Fonseca PLC, Tomé LMR, Vaz ABM, Badotti F, Ramos RTJ, Brenig B: Draft genome sequence of Trametes villosa (Sw.) Kreisel CCMB561, a tropical white-rot Basidiomycota from the semiarid region of Brazil. *Data in brief* 2018, 18:1581-1587.

5. Knežević A, Stajić M, Sofrenić I, Stanojković T, Milovanović Ī, Tešević V, Vukojević J: Antioxidative, antifungal, cytotoxic and antineurodegenerative activity of selected Trametes species from Serbia. *PloS one* 2018, 13(8):e0203064.

6. Yang XQ, Zhao XX, Liu CY, Zheng Y, Qian SJ: Decolorization of azo, triphenylmethane and anthraquinone dyes by a newly isolated Trametes sp. SQ01 and its laccase. *Process Biochemistry* 2009, 44(10):1185-1189.

7. Levin L, Herrmann C, Papinutti VL: Optimization of lignocellulolytic enzyme production by the white-
rot fungus Trametes trogii in solid-state fermentation using response surface methodology. 

Biochemical Engineering Journal 2008, 39(1):207-214.

8. Neves M, Baseia I, Drechsler-Santos E, Góes-Neto A: Guide to the common fungi of the semiarid region of Brazil. TECC Editora, Florianópolis 2013.

9. Milton RD, Giroud F, Thumser AE, Minteer SD, Slade RC: Hydrogen peroxide produced by glucose oxidase affects the performance of laccase cathodes in glucose/oxygen fuel cells: FAD-dependent glucose dehydrogenase as a replacement. Physical Chemistry Chemical Physics 2013, 15(44):19371-19379.

10. Salaj-Kosla U, Pöller S, Schuhmann W, Shleev S, Magner E: Direct electron transfer of Trametes hirsuta laccase adsorbed at unmodified nanoporous gold electrodes. Bioelectrochemistry 2013, 91:15-20.

11. Davies GJ, Williams SJ: Carbohydrate-active enzymes: sequences, shapes, contortions and cells. Biochemical Society Transactions 2016, 44(1):79-87.

12. Sista Kameshwar AK, Qin W: Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. Mycology 2018, 9(2):93-105.

13. Zhao Z, Liu H, Wang C, Xu J-R: Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC genomics 2013, 14(1):274.

14. Dai W, Chen X, Wang X, Xu Z, Gao X, Jiang C, Deng R, Han G: The Algicidal Fungus Trametes versicolor F21a Eliminating Blue Algae via Genes Encoding Degradation Enzymes and Metabolic Pathways Revealed by Transcriptomic Analysis. Front Microbiol 2018, 9:826.

15. Newman DJ, Cragg GM: Natural products as sources of new drugs over the 30 years from 1981 to 2010. Journal of natural products 2012, 75(3):311-335.

16. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, De Los Santos EL, Kim HU, Nave M: antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic acids research 2017, 45(W1):W36-W41.

17. Pusztahelyi T, Holb IJ, Pocsi I: Secondary metabolites in fungus-plant interactions. Frontiers in plant science 2015, 6:573.
18. Keller NP: Fungal secondary metabolism: regulation, function and drug discovery. *Nature reviews Microbiology* 2018.

19. Collemare J, Billard A, Böhnert HU, Lebrun M-H: Biosynthesis of secondary metabolites in the rice blast fungus Magnaporthe grisea: the role of hybrid PKS-NRPS in pathogenicity. *Mycological research* 2008, 112(2):207-215.

20. Cimermancic P, Medema MH, Claesen J, Kurita K, Brown LCW, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J: Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 2014, 158(2):412-421.

21. Clevenger KD, Bok JW, Ye R, Miley GP, Verdan MH, Velk T, Chen C, Yang K, Robey MT, Gao P: A scalable platform to identify fungal secondary metabolites and their gene clusters. *Nature chemical biology* 2017, 13(8):895.

22. Nielsen JC, Grijseels S, Prigent S, Ji B, Dainat J, Nielsen KF, Frisvad JC, Workman M, Nielsen J: Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in Penicillium species. *Nat Microbiol* 2017, 2:17044.

23. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW: Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences* 2012, 109(16):6241-6246.

24. Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H: ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular ecology resources* 2013, 13(2):218-224.

25. de Oliveira Carneiro RT, Lopes MA, Silva MLC, da Silva Santos V, De Souza VB, De Sousa AO, Pirovani CP, Koblitz MGB, Benevides RG, Goes-Neto A: Trametes villosa lignin peroxidase (TvLiP): genetic and molecular characterization. *Journal of microbiology and biotechnology* 2017, 27(1):179-188.

26. Berrin JG, Navarro D, Couturier M, Olive C, Grisel S, Haon M, Taussac S, Lechat C, Courtecuisse R, Favel A et al: Exploring the natural fungal biodiversity of tropical and temperate forests toward improvement of biomass conversion. *Applied and environmental microbiology* 2012, 78(18):6483-
27. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic acids research* 2008, 37(suppl_1):D233-D238.

28. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y: dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic acids research* 2012, 40(W1):W445-W451.

29. Murphy C, Powlowski J, Wu M, Butler G, Tsang A: Curation of characterized glycoside hydrolases of fungal origin. *Database* 2011, 2011.

30. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010, 11:119.

31. Aherfi S, Andreani J, Baptiste E, Oumessoum A, Dornas FP, dos SP A, Claudia A, Chabriere E, Abrahao J, Levasseur A: A large open pangenome and a small core genome for giant pandoraviruses. *Frontiers in microbiology* 2018, 9:1486.

32. Li L, Stoeckert CJ, Jr., Roos DS: OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome research* 2003, 13(9):2178-2189.

33. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, de Sousa F: Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in ecology and evolution* 2013, 4(10):914-919.

34. Lesage-Meessen L, Haon M, Uzan E, Levasseur A, Piumi F, Navarro D, Taussac S, Favel A, Lomascolo A: Phylogeographic relationships in the polypore fungus Pycnoporus inferred from molecular data. *FEMS microbiology letters* 2011, 325(1):37-48.

35. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 2004, 32(5):1792-1797.

36. Talavera G, Castresana J: Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007, 56(4):564-577.

37. JD R: Phylogenetic analysis using PHYLIP. *Methods Mol Biol* 1999, 132(132):243-258.
38. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 2011, 28(10):2731-2739.

39. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN: The COG database: an updated version includes eukaryotes. *BMC bioinformatics* 2003, 4(1):41.

40. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M: eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic acids research* 2015, 44(D1):D286-D293.

41. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014, 42(D1):D490-D495.

42. Mäkelä M, DiFalco M, McDonnell E, Nguyen TTM, Wiebenga A, Hildén K, Peng M, Grigoriev I, Tsang A, de Vries R: Genomic and exoproteomic diversity in plant biomass degradation approaches among Aspergilli. *Studies in Mycology* 2018, 91:79-99.

43. Lomascolo A, Cayol J-L, Roche M, Lin G, Robert J-L, Record E, Lesage-Meessen L, Ollivier B, Sigoillot J-C, Asther M: Molecular clustering of Pycnoporus strains from various geographic origins and isolation of monokaryotic strains for laccase hyperproduction. *Mycological Research* 2002, 106(10):1193-1203.

44. Levasseur A, Lomascolo A, Chabrol O, Ruiz-Dueñas FJ, Boukhris-Uzan E, Piumi F, Kües U, Ram AF, Murat C, Haon M: The genome of the white-rot fungus Pycnoporus cinnabarinus: a basidiomycete model with a versatile arsenal for lignocellulosic biomass breakdown. *Bmc Genomics* 2014, 15(1):486.

45. Busk PK, Lange M, Pilgaard B, Lange L: Several genes encoding enzymes with the same activity are necessary for aerobic fungal degradation of cellulose in nature. *PLoS One* 2014, 9(12):e114138.

46. Couturier M, Navarro D, Chevret D, Henrissat B, Piumi F, Ruiz-Dueñas FJ, Martinez AT, Grigoriev IV, Riley R, Lipzen A: Enhanced degradation of softwood versus hardwood by the white-rot fungus Pycnoporus coccineus. *Biotechnology for biofuels* 2015, 8(1):216.
47. Pavlov AR, Tyazhelova TV, Moiseenko KV, Vasina DV, Mosunova OV, Fedorova TV, Maloshenok LG, Landesman EO, Bruskin SA, Psurtseva NV: Draft genome sequence of the fungus Trametes hirsuta. *Genome announcements* 2015, 3(6):e01287-01215.

48. Cerrón LM, Romero-Suárez D, Vera N, Ludeña Y, Villena GK, Gutiérrez-Correa M: Decolorization of textile Reactive dyes and effluents by biofilms of Trametes polyzona LMB-TM5 and Ceriporia sp. LMB-TM1 isolated from the Peruvian Rainforest. *Water, Air, & Soil Pollution* 2015, 226(8):235.

49. Granchi Z, Peng M, Chi-A-Woeng T, de Vries RP, Hildén K, Mäkelä MR: Genome Sequence of the Basidiomycete White-Rot Fungus Trametes pubescens FBCC735. *Genome announcements* 2017, 5(8):e01643-01616.

50. Wang J, Zhang Y, Xu Y, Fang W, Wang X, Fang Z, Xiao Y: Genome sequence of a laccase producing fungus Trametes sp. AH28-2. *Journal of biotechnology* 2015, 216:167-168.

51. Floudas D, Binder M, Riley R, Barry K, Blanchette R, Henrissat B, Martínez A, Otillar R, Spatafora J, Yadav J: The Paleozoic origin of enzymatic mechanisms for decay of lignin reconstructed using 31 fungal genomes. *Science* 2012, 336(6089):1715-1719.

Table

Table 1 Comparison of the genomic features of 9 strains in *Trametes* genus.

| Organism         | DB accession number | Contigs/Scaffolds | Genome size(bp) | CDSs/ORFs* | Reference |
|------------------|---------------------|-------------------|-----------------|------------|-----------|
| *T. cinnabarina* | BRFM137             | 776               | 33,668,918      | 59,163     | [43, 44]  |
| *T. cinnabarina* | JSYY01000001        | 9,648             | 31,621,284      | 61,560     | [45]      |
| *T. coccinea*    | KZ084086            | 222               | 32,757,859      | 56,735     | [46]      |
| *T. hirsuta*     | CP019371            | 13                | 37,433,966      | 61,150     | [47]      |
| *T. polyzona*    | MNAD01000001        | 5,321             | 36,629,716      | 65,446     | [48]      |
| *T. pubescens*   | NLJ01000001         | 1,731             | 39,739,821      | 67,916     | [49]      |
| *T. sp.AH28-2*   | NW_007360321        | 306               | 38,904,488      | 65,191     | [50]      |
| *T. versicolor*  | PUDQ01000001        | 283               | 44,794,008      | 71,443     | [51]      |

Note: * represent the CDSs/ORFs were predicted by Pordigal.

Figures
Figure 1

Genetic diversity of strains in Trametes genus. a. The size of pan-genome (red) and core genome (blue) shared by different genomes, respectively. b. Core and specific gene families of 9 strains of Trametes genus. The number core genome shared by all strains is in the center (8,518). The number of non-overlapping portions of each oval represents the size of specific families. c. The phylogenetic tree based on the ITS1 sequences of 9 strains of Trametes genus and Coriolopsis caperata. d. The phylogenetic tree based on the ITS2 sequences of 9 strains of Trametes genus and Coriolopsis caperata. C. caperata was selected as outgroup to root the topology of phylogenetic tree. e. The UPGMA tree of Manhattan distance based on the pan-genome composition of 9 strains of Trametes genus.
Figure 2

COG and GO annotation of gene families. a. The distribution of COG annotation of pan-genome that assigned in 25 functional categories. b. The distribution of GO annotation of pan-genome in biological process, molecular function and cellular component. Accessory clusters (pink), core clusters (green), and specific genes (wathet).
The distributions of CAZymes in 9 Trametes strains of Trametes genus. a. The distribution of CAZymes in pan-genome of Trametes genus. b. The distribution of CAZymes of pan-genome in each Trametes strain grouped by accessory clusters, core clusters and specific genes. The numbers shown in subgraph represents the orthologous genes assigned by CAZymes and grouped by accessory clusters (pink), core clusters (green), and specific genes (wathet). c. The distribution of CAZymes of accessory clusters in each Trametes strain. d. The distribution of CAZymes of core clusters in each Trametes strain. e. The distribution of CAZymes of specific genes in each Trametes strain. The numbers shown in subgraphs represent the orthologous genes assigned by CAZymes. GT, glycosyltransferase; GH, glycoside hydrolase; CE, carbohydrate esterase; CBM, carbohydrate-binding molecule; AA, auxiliary activities; PL, polysaccharide lyases.
The distribution of biosynthetic gene clusters in 9 Trametes strains of Trametes genus. The distribution of biosynthetic gene clusters in each Trametes strain. The numbers shown in graph represent the counts of different types of biosynthetic gene clusters predicted by antiSMASH for each Trametes strain. Note: cf_putative, putative cluster of unknown type identified with the ClusterFinder algorithm; terpene, terpene cluster; nrps, nonribosomal peptide synthetase cluster; cf_fatty_acid, putative fatty acid cluster identified with the ClusterFinder algorithm; t1pks, type I PKS cluster; lantipeptide, lanthipeptide cluster; other, cluster containing a secondary metabolite-related protein that does not fit into any other category.

**a** MNAD01001604.1-Gene Cluster 13. Type = nrps. ClusterFinder probability: 0.8147

| cf_putative | terpene | nrps | cf_fatty_acid | t1pks | lantipeptide | other |
|-------------|---------|------|--------------|-------|-------------|-------|
| 3           | 9       | 1    | 0            | 0     | 0           | 5     |
| 24          | 12      | 1    | 2            | 5     | 0           | 8     |
| 18          | 9       | 1    | 1            | 1     | 0           | 7     |
| 21          | 10      | 1    | 1            | 1     | 0           | 5     |
| 21          | 11      | 2    | 1            | 2     | 1           | 7     |
| 39          | 12      | 1    | 1            | 4     | 1           | 8     |
| 34          | 17      | 2    | 1            | 1     | 0           | 6     |
| 31          | 11      | 1    | 1            | 1     | 0           | 8     |
| 35          | 11      | 2    | 1            | 1     | 1           | 8     |

**b** MNAD01001612.1-Gene Cluster 15. Type = other. ClusterFinder probability: 0.8363

| cf_putative | terpene | nrps | cf_fatty_acid | t1pks | lantipeptide | other |
|-------------|---------|------|--------------|-------|-------------|-------|
| CCBP010000287_c1: Trametes cinnabaria BRFM137 (66% of genes show similarity) |
| JH711791_c1: Trametes versicolor FP-101664 SS1 (66% of genes show similarity) |
| JH719402_c1: Dichomitus squalens LYAD-421 SS1 (53% of genes show similarity) |
| HE797138_c1: Fibroporia radiculosa TFFH 294 (46% of genes show similarity) |
| KB445802_c1: Ceriporiopsis subvermispora B (46% of genes show similarity) |
The probability and homologous gene clusters for three putative BGCs of *T. pubescens*. The probability and homologous gene clusters for the putative BGCs of a: No. 13, b: No. 15 and c: No. 20. In the subgraph, the same colors of the boxes with arrows are similar genes.