Comparative genomic analysis of five Coprinus species

Jingjing Wang
Hefei Normal University

Ran Zhang
Anhui University

Guoao Ding
Anhui University

Lingling Wang
Hefei Normal University

Wei Wang
Hefei Normal University

Yan Zhang (✉ zhangyanhfnu@gmail.com)
Hefei Normal University

GuiLan Zhu
Hefei Normal University

Research Article

Keywords: fungal laccase, Coprinus species, comparative genomic analysis, gene family, average nucleotide identity, positive selection

Posted Date: January 12th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2166905/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Although *Coprinus* is widely known for the phenomenon of deliquescence and production of fungal laccases and extracellular peroxygenases, the genome structure and genetic diversity of *Coprinus* species have not been extensively studied. To reveal the genomic structure and diversity in *Coprinus* species, the genomes of five *Coprinus* species were compared and analyzed. A total of 24,303 orthologous gene families, including 89,462 genes, were identified in the five species. The numbers of core, softcore, dispensable, and private genes were 5617 (25.6%), 1628 (7.4%), 2083 (9.5%), and 12574 (57.4%), respectively. Differentiation time analysis revealed that *Coprinellus micaceus* and *Coprinellus angulatus* differentiated approximately 181.0 million years ago. *Coprinopsis cinerea* and *Coprinopsis marcescibilis* differentiated approximately 131.0 million years ago, and they were differentiated from *Candolleomyces aberdarensis* approximately 176.0 million years ago. Gene family contraction and expansion analyses showed that 1,465 genes and 532 gene families were expanded, and 95 genes and 134 gene families were contracted. Ninety-five laccase-coding genes were detected in the five species, and the distribution of the laccase-coding genes in the five species was not uniform. These data provide a reference for a deeper understanding of the genetic structure of the genomes of *Coprinus* species. Furthermore, this study provides a reference for follow-up studies on the genome structure of *Coprinus* species and the diversity of specific functional genes.

Background

*Coprinus* species are cosmopolitan (.Padamsee et al. 2008), and widely renowned for a phenomenon called deliquescence (.Nagy et al. 2010). During this process, fruiting body tissues become blackish inky fluid by autodigestion of the fruiting body cells upon maturation (.Hopple and Vilgalys 1999). Another notable feature of *Coprinus* is that most species can produce fungal laccases(.Hoegger et al. 2004, Hu et al. 2019, Kilaru et al. 2006, Lin et al. 2013) and extracellular peroxygenases (.Aranda et al. 2009).

Laccases are biologically important enzymes that belong to the oxidase group and are useful as green enzymes for cleaner industrial applications to reduce environmental pollution (.Senthivelan et al. 2016).

Considering that these enzymes have various substrate catalytic properties and numerous applications in various fields, they have received attention from researchers for their use in further intensive studies worldwide (.Senthivelan, Kanagaraj and Panda 2016).

Although laccases are produced by plants, bacteria, insects, and fungi(.Dwivedi et al. 2011, Senthivelan, Kanagaraj and Panda 2016), fungal laccases have been the most extensively studied (.Liu et al. 2022, Mougin et al. 2003). However, the distribution of deliquescence and laccase- and peroxygenase-coding genes in *Coprinus* species does not exhibit a clear pattern, which may be due to the unclear classification system of *Coprinus* and the relatively complex genomes of *Coprinus* species (.Nagy, Urban, Örstadius, Papp, Larsson and Vágvölgyi 2010, Örstadius et al. 2015).

Traditionally, Psathyrellaceae species have been classified into two large genera, *Coprinus* and *Psathyrella*. However, the two genera have many common features, including similar habitat preferences, spore characteristics, degradation of spore pigments in sulfuric acid, and developmental, morphological,
and ecological traits that are significantly convergent; several alternative classifications have also been proposed (Nagy, Urban, Örstadius, Papp, Larsson and Vágvölgyi 2010). Furthermore, it has long been recognized that deliquescent taxa within Psathyrellaceae do not form a monophyletic group (Vašutová et al. 2008, Walther et al. 2005). The genus Coprinus was proposed to be split into four genera (Redhead et al. 2001). Although some studies have added valuable information to our knowledge of the phylogeny of Psathyrellaceae (Nagy, Urban, Örstadius, Papp, Larsson and Vágvölgyi 2010, Örstadius, Ryberg and Larsson 2015, Wächter and Melzer 2020). However, these studies were all based on the diversity analysis of one or several limited genes. Even though these studies provide substantial guidance for the study of the phylogeny of Psathyrellaceae, they fail to provide effective information to study the origin and maintenance mechanisms of Psathyrellaceae functional genes.

Comparative genomics provide an important technical means for studying the origin and maintenance of fungal genetic diversity (de Vries et al. 2017, Kiss et al. 2019, Ma et al. 2013, Zhang et al. 2020). However, the genome structure and genetic diversity of Coprinus species have not been extensively studied. Therefore, in this study, we aimed to reveal the genomic structure and genetic diversity of Coprinus species by comparing and analyzing the genomes of five Coprinus species, i.e. Coprinellus angulatus, Coprinellus micaceus, Coprinopsis cinerea, Coprinopsis marcescibilis, and Candolleomyces aberdarensis.

Materials And Methods

Genomic data collection of Coprinus species

The published genomes of Coprinus species were searched for and retrieved from NCBI, and the following five fully annotated fungi of the genus were retrieved: C. aberdarensis (60.61 Mb), C. angulatus (59.3 Mb), C. micaceus (77.39 Mb), C. cinerea (36.19 Mb), and C. marcescibilis bilis (38.91Mb). T. mesenterica was selected as the outgroup (Table S1).

Gene family clustering and enrichment

Based on the amino acid sequences, gene family clustering was performed using OrthoFinder version 2.3.12 (parameter -M msa) (Emms and Kelly 2019), in which blastp version 2.6.0 (parameter -evalue 1e-5 -outfmt 6) (Camacho et al. 2009) was used for alignment. The results are shown using the R clusterProfiler package (Wu et al. 2021) based on the functional annotation results of GO (Ashburner et al. 2000) and Kyoto encyclopedia of genes and Genomes (KEGG) (Kanehisa and Goto 2000). The numbers of non-redundant core, softcore, dispensable, and private gene families in each species and all species were counted.

Analysis Core and private gene families

A Venn diagram was drawn using the Perl script. Genes common to all genomes were defined as core genes, those common to 90% or more genomes were defined as softcore genes, those private to each
genome were defined as private genes, and the remaining genes were defined as dispensable genes. Core and private gene analyses were performed using the R clusterProfiler package (.Wu, Hu, Xu, Chen, Guo, Dai, Feng, Zhou, Tang, Zhan, Fu, Liu, Bo and Yu 2021) based on the private gene and functional annotation results of GO (.Ashburner, Ball, Blake, Botstein, Butler, Cherry, Davis, Dolinski, Dwight, Eppig, Harris, Hill, Issel-Tarver, Kasarskis, Lewis, Matsese, Richardson, Ringwald, Rubin and Sherlock 2000) and KEGG (.Kanehisa and Goto 2000).

**Phylogeny analysis and calculation of differentiation time**

Multiple sequence alignment was performed on the protein sequences of each single copy gene using MUSCLE version 3.8.31 (.Edgar 2004), and then the alignment results were filtered using trimAl v1.4. rev22 (parameter -gt 0.2) (.Capella-Gutiérrez et al. 2009), and then the filtered alignment sequences were merged and connected to the supergenes. Finally, an ML phylogenetic tree was constructed based on the supergenes using RAxML version 8.2.10 (.Stamatakis 2014) with the PROTGAMMAWAG model.

The fossil time can make the calculation result of the differentiation time more accurate. Fossil timetables were obtained from TIMETREE (http://www.timetree.org/). Based on the topological structure of the phylogenetic tree and the fossil timetable, the differentiation times of the species were estimated using the mcmtree subprogram (parameters nsample = 3000000; burnin = 8000000; seqtype = 0; model = 4) of PAML version 4.9 (.Yang 2007).

**Gene family contraction and expansion**

The number of gene family members of the ancestors of each branch was estimated using the birth-mortality model based on the species evolutionary tree and gene family clustering results through café version 3.1 (.Han et al. 2013), thereby the contraction and expansion of the gene family of the species relative to the ancestors were predicted.

**Analysis of laccase synthesis gene family**

Laccase gene family numbers were searched on the InterPro website (https://www.ebi.ac.uk/interpro/result/InterProScan/), and PF numbers PF00394, PF07731, and PF07732 were obtained. Gene IDs were searched according to the PF numbers, and protein sequences were obtained according to the gene IDs. An ML phylogenetic tree based on laccase protein sequences was constructed as described above. Motif locations in laccase protein sequences were identified using MEME version 5.4.1 (.Bailey and Elkan 1994).

**Genome average nucleotide identity analysis**

ANI indicates the similarity of all orthologous protein-coding genes between two genomes and is often used to indicate the evolutionary distance between genomes (.Pritchard et al. 2015). ANI was calculated using the pyani ANIm algorithm (https://pureportal.strath.ac.uk/en/publications/pyani-v028-average-nucleotide-identity-ani-and-related-measures-f).
Results

Gene family clustering and enrichment

Five available genomic datasets from five *Coprinus* species were collected for comparative genome analysis. The genome size of the five species ranged from 36.19 to 77.39 Mb. A total of 21902 orthologous gene families were identified, including 89,462 genes (Fig. 1A). The number of single-copy gene families was relatively stable in different species, with an average copy number of 3134.8 ± 58.98 (Fig. 1A). In this study, 5,668 species-specific gene families were identified. Notably, in *C. micaceus*, the number of paralogous genes was markedly higher than in other species, reaching 5,672 private genes (Fig. 1A and 1B). The numbers of core, softcore, dispensable, and private genes were 5617 (25.6%), 1628 (7.4%), 2083 (9.5%), and 12574 (57.4%), respectively (Fig. 1C and Table S2). Kyoto ontology (KO) and gene ontology (GO) enrichment results showed that the core genes were primarily involved in regulating energy metabolism, biomass synthesis, and metabolic processes, such as translation, mitochondrion, ribosome, rRNA processing, oxidative phosphorylation, biosynthesis amino acids, and fatty acid metabolism (Fig. 2). However, the private genes were primarily involved in some kinase, superoxide dismutase, monooxygenase, oxidoreductase, and peroxidase activities, structural constituents of the cell wall, and response to oxidative stress, which endowed fungi with the ability to participate in metabolism of substances such as tyrosine, glutathione, sphingolipid, glyoxylate, and dicarboxylate, and adapt to different habitats (Fig. 3).

Phylogenetic tree, differentiation time, and gene family contraction and expansion

The phylogenetic tree constructed using 2,353 single-copy genes showed that *C. micaceus* and *C. angulatus* were clustered into one branch, and then clustered with *C. aberdarensis*. *C. cinerea* and *C. marcescibilis* were clustered into one branch. *Tremella mesenterica* was furthest from its evolution (Fig. 4A). Differentiation time analysis showed that *C. micaceus* and *C. angulatus* differentiated approximately 181.0 million years ago. *C. cinerea* and *C. marcescibilis* differentiated approximately 131.0 million years ago, and they have differentiated from *C. aberdarensis* approximately 176.0 million years ago (Fig. 4B). However, genome average nucleotide identity (ANI) analysis showed that, as an outgroup, the ANI between *T. mesenterica* and *C. marcescibilis* was 1, and that between *T. mesenterica* and *C. angulatus* was 0.93, which was higher than that between *C. marcescibilis* and *C. cinerea*, as well as that between *C. micaceus* and *C. angulatus* (0.84; Table S3).

Gene family contraction and expansion analyses showed that 1,465 genes and 532 gene families were expanded, and 95 genes and 134 gene families were contracted (Fig. 4C and Table S4). Furthermore, our results also indicated that the expansion and contraction gene families in the *C. micaceus* genome were 532 and 134, respectively, in which the expansion gene families were more numerous than the contraction gene families. However, the expansion and contraction gene families in *C. aberdarensis* were 301 and 722, respectively, in which the expansion gene families were fewer than the contraction gene
families (Fig. 4C). Compared with *Coprinopsis*, the genomes of the two *Coprinellus* species contained more expansion and contraction gene families (Fig. 4C).

The GO and KO enrichment results showed that the expansion genes mainly participated in carbohydrate metabolic processes, catalytic activity, and catabolic processes, which endowed the fungi with the ability to synthesize and metabolize substances and adapt to different habitats (Fig. 5). The contraction genes exhibited oxidoreductase activity and participated in protein processing in the endoplasmic reticulum and export (Fig. 5).

**Laccase gene family**

Ninety-five laccase genes were detected in the five species. The phylogenetic tree of laccase genes showed that these genes were highly diverse (Fig. 6). Most laccases contain 10 motifs in a similar order (Fig. 6). However, the distribution of laccase genes among the five species was not uniform. The *C. micaceus* genome contained the most multicopper oxidase PF07731 genes, the *C. marcescibilis* genome contained the most multicopper oxidase PF00394 genes, and the *C. aberdarensis* genome contained the most multicopper oxidase PF07732 genes (Fig. S1).

**Discussion**

With the in-depth development of sequencing technology and bioinformatics analysis tools, comparative genomics has been widely used in studies on species evolution and formation, ecological diffusion, habitat adaptation, and antibiotic resistance gene diffusion (Lin et al. 2018, Ni 2016, Qin et al. 2021, Shah et al. 2022, Wang et al. 2022, Yao et al. 2022). The core genome is generally considered as the minimum genome necessary for the survival of free-living organisms (Pandaranayaka et al. 2019, Zafar et al. 2002). Therefore, core genes are closely related to the metabolic processes necessary for survival. Private genes are usually related to the secondary metabolites of fungi and enable different fungi to synthesize different small molecules and adapt to different habitats (Keller 2019, Pandaranayaka, Frenkel, Elad, Prusky and Harel 2019, Slot and Gluck-Thaler 2019, Wisecaver et al. 2014). Our results indicate that the core genes were mainly involved in the regulation of energy metabolism, biomass synthesis, and metabolic processes, whereas the private genes mainly involved some kinase, superoxide dismutase, monoxygenase, oxidoreductase, and peroxidase activities, structural constituents of the cell wall, and response to oxidative stress, which endowed fungi with the ability to participate in substance metabolism, such as tyrosine, glutathione, sphingolipid, glyoxylate, and dicarboxylate, and adapt to different habitats. These results indicate that private genes endow different fungi with different secondary metabolic characteristics and habitat adaptability, which is of substantial value for maintaining fungal biodiversity. Notably, the expansion genes mainly participated in carbohydrate metabolic processes, catalytic activity, and catabolic processes, which endowed the fungi with the ability to synthesize and metabolize secondary metabolites and adapt to different habitats. These results imply that fungal differentiation is primarily due to the differentiation of private genes. Considering that secondary metabolites are mainly produced by private genes, it may be of considerable significance to use secondary metabolites of fungi as a basis for fungal classification.
As sequencing technology has advanced, fungal classification has shifted from morphological characteristics to molecular phylogenetic evidence (Nagy, Urban, Örstadius, Papp, Larsson and Vágvölgyi 2010, Örstadius, Ryberg and Larsson 2015). Recently, Wächter and Melzer (2020) inferred that the family Psathyrellaceae forms distinct phylogenetic clades and is divided into 16 genera based on molecular phylogenetic evidence and morphological characteristics (Wächter and Melzer 2020). Although some morphological features caused by private genes may be used as the basis for fungal classification, the secondary metabolites produced by private genes are difficult to reflect in morphology, so it is difficult to use them for fungal morphological classification. Furthermore, genome ANI analysis showed that, as an outgroup, the ANI between T. mesenterica and C. marcescibilis was 1 and that between T. mesenterica and C. angulatus was 0.93, which was higher than that between C. marcescibilis and C. cinerea, as well as that between C. micaceus and C. angulatus. These results indicate that it is difficult to classify fungi effectively using only a few genes. Therefore, it is necessary to identify secondary metabolites for fungal classification.

Conclusions

A total of 24,303 orthologous gene families were identified in the five species, including 89,462 genes. The core genes were mainly involved in the regulation of energy metabolism, biomass synthesis, and metabolic processes, whereas the private genes endowed the fungi with the ability to participate in substance metabolism and adapt to different habitats. These results indicate that private genes endow different fungi with different secondary metabolic characteristics and habitat adaptability, which is of considerable importance for maintaining fungal biodiversity. These results imply that fungal differentiation is mainly due to the differentiation of private genes. Considering that secondary metabolites are primarily produced by private genes, using secondary metabolites of fungi as a basis for fungal classification may be of great significance.

Declarations

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Competing interests

The authors declare no conflict of interest.

Funding
This study was funded by the National Natural Science Foundation of China, grant number 31800049; Natural Science Foundation of Higher Education in Anhui Province, grant number KJ2021ZD0113; the Program for Young Outstanding Talents in Anhui Province, grant number gxyqZD2022071; the Opening Laboratory Programme for the Medicinal and Food Homologous Natural Resources Exploration in Anhui Province, grant number YSTY2022028&YSTY2022030.

Authors' contributions

Yan Zhang and Jingjing Wang designed the experiments. Ran Zhang, Guoao Ding, Wei Wang and Lingling Wang performed the experiments. Yan Zhang and GuiLan Zhu collected and analyzed the data. Yan Zhang and GuiLan Zhu wrote the draft of the manuscript. All authors revised and approved the final version of the manuscript.

Acknowledgements

Thanks Jiajia Ni, a technician at Guangdong Meilikang Bio-Science Ltd., China for assistance with data analysis.

References


Figures

(A) The number of homologous genes (A), core and private genes (B), and the number of genes of each classification (C).

Figure 1

The number of homologous genes (A), core and private genes (B), and the number of genes of each classification (C).
Figure 2

Gene ontology (A) and Kyoto ontology (B) enrichment of core genes
Figure 3

Gene ontology (A) and Kyoto ontology (B) enrichment of private genes
Figure 4

Figure 5

GO and KO enrichment of expansion and contraction gene families. (A)–(D) shows the GO and KO enrichment of expansion gene families, and GO and KO enrichment of contraction gene families, respectively.

Figure 6

Phylogenetic tree of laccase genes and motif distribution in the laccases. The motif sites were predicted by MEME plus any additional sites detected using a motif scanning algorithm. The MEME sites are shown in solid color and additional scanned sites are shown in transparent color. Hovering the cursor
over a site will reveal details about the site. Only sequences containing a predicted or scanned motif site are shown. The scanned sites are predicted using a log-odds scoring matrix constructed from the MEME sites. Only scanned sites with position p-values less than 0.0001 are shown.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- ZhangY2.2manuscriptSM.docx