

RESEARCH

Open Access



Whole genome sequencing and analysis of the symbiotic *Armillaria gallica* M3 with *Gastrodia elata*

Cheng-Ying Luo^{1,2†}, Yao Lu^{1,3†}, Lei Su^{1,3}, Jia-Jia Liu^{1,3}, Jia-Yun Miao⁴, Yi-Cen Lin^{1,3*} and Lian-Bing Lin^{1,3*}

Abstract

Background On the one hand, *Armillaria* is regarded as a plant disease that causes serious root rot of forest trees, on the other hand, *Armillaria* is also an important symbiotic fungi of the valuable Chinese herb *Gastrodia elata*. Currently, the whole genome database of *Armillaria* is relatively limited, and it is expected that a more comprehensive understanding of the symbiotic interactions between *Armillaria* and *G. elata* can be achieved through genome-wide comparisons and functional annotations. Whole genome sequencing of *Armillaria gallica* M3 strain was performed using Oxford Nanopore Technologies sequencing platform, and the sequencing data were used to perform genome assembly, gene prediction and functional annotation, carbohydrate-active enzymes, and host–pathogen interactions using bioinformatics methods.

Results In this study, we obtained an 83.33 M genome of *A. gallica* M3 strain, which consisted of 38 overlapping clusters with an N50 of 6,065,498 bp and a GC content of 47.43%. A total of 12,557 genes were identified in the genome of *A. gallica* M3, and the repetitive sequences accounted for about 44.36% of the genome. 42.26% of the genome was composed of glycoside hydrolases (GHs), 16.15% of the genome was composed of glycosyltransferases (GTs). In addition, 3412 genes in *A. gallica* M3 were involved in the host–pathogen interaction mechanism.

Conclusions These results have elucidated the characteristics of *A. gallica* M3 from a genomic perspective to a certain extent. They help to analyze the inner mechanism of *A. gallica* M3 being able to symbiosis with *G. elata* at the genomic level, which is of great significance to the next related research of *A. gallica* M3.

Keywords *Armillaria*, *Gastrodia elata*, Symbiosis, Genome, Mechanism analysis

Introduction

Gastrodia elata Bl., a perennial herb of the *Gastrodia* in the Orchidaceae, has been used in medicine for more than 2,000 years, and is a traditional and valuable Chinese medicine in China [1]. *G. elata* has anticonvulsant, hypnotic, sedative, antidepressant effects and other pharmacological effects [2]. *G. elata* to dried rhizomes into medicine, the main medicinal ingredients for gastrodin [3]. Since *G. elata* has no roots or leaves and no chloroplasts, it cannot carry out photosynthesis. The growth process of *G. elata* to *Armillaria* as the only source of nutrients, through the digestion of *Armillaria* invasion of rhizomorphs to obtain nutrients [4, 5]. And when

[†]Cheng-Ying Luo and Yao Lu contributed equally to this work.

*Correspondence:

Yi-Cen Lin

yicenlin@kust.edu.cn

Lian-Bing Lin

linlb@kust.edu.cn

¹ Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming 650500, China

² Gastrodia Industry Development Center of Yiliang County, Zhaotong 657000, China

³ Engineering Research Center for Replacement Technology of Feed Antibiotics of Yunnan College, Kunming 650500, China

⁴ Yunnan Senhao Fungi Industry Co., Ltd, Zhaotong 657600, China



the *Armillaria* source of nutrition is insufficient, it will use the nutrients in the body of *G. elata* for its growth. Therefore, *Armillaria* and *G. elata* to establish a special symbiotic relationship [6].

Most terrestrial plants form symbiotic relationships with fungi, and these mycorrhizal fungi can influence soil structure and regulate soil nutrient and carbon cycling [7]. *Armillaria* are one of the most globally important genera of fungal root pathogens and one of the largest and oldest organisms on earth [8, 9]. The ability of *Armillaria* to act as saprophytes, especially white rot fungi, is considered to be beneficial to natural ecosystems because they degrade lignin and therefore play an important role in the carbon cycle [8].

Armillaria gallica, an important plant pathogenic fungus, often causes root rot in many trees. However, it is extremely special that *Armillaria* is an essential fungus in the growth process of *G. elata*, which can provide nutrients for *G. elata*. At present, there are more than 70 known species of *Armillaria*, but only a few of them can be symbiotic with *G. elata* [10]. As the ability of different strains of *Armillaria* to form rhizomorphs and the ability of rhizomorphs to infest the cortical tissue of *G. elata* differs, the quality of *Armillaria* will directly affect the yield of *G. elata* [11]. Liu et al. found that the extracellular secretion of pectinase, xylanase, cellulase, laccase and other enzymes not only provides nutrients for the growth and development of *A. gallica*, but also provides a material basis for *A. gallica* to infest the epidermis of *G. elata* [12]. In other words, the activity of extracellular enzymes of *A. gallica* is closely related to the yield and quality of *G. elata*. Meanwhile, Zhao et al. showed that the extracellular enzyme activity was stronger and polysaccharide content was higher in the *A. gallica* strain [13]. Moreover, *A. gallica* has strong parasitic ability, and its mycelium is thick and developed, which are beneficial to the cultivation and production of *G. elata*. [3]. *A. gallica* polysaccharides have a variety of pharmacological activities, including antioxidant effects, anti-fatigue and anti-inflammatory effects [14–16].

Compared to most other seed plants, *G. elata*'s mitochondrial genome significantly enlarged [1]. It was also found that the number of monocotyledonous mannose-binding lectin *Gastrodia* antifungal protein (GAFP) genes was increased in *G. elata*, and the GAFP protein could inhibit the growth of ascomycete and basidiomycete, including *Armillaria* and *Ganoderma lucidum*, among others [17, 18]. More than 80% of the GAFP genes expressed highly in protocorms and juvenile tubers before *G. elata* established a stable symbiotic relationship with *Armillaria* [1]. Strigolactone promotes *Armillaria* mycelial branching and contributes to the establishment of a symbiotic relationship between *G. elata* and

Armillaria [19]. In *G. elata*, the number of strigolactone biosynthesis and transport key genes carotenoid cleavage dioxygenases (CCDs) and ABC transporters was increased; the number of calmodulin-dependent protein kinase genes of the *does-not-make-infections* 3 subfamily (*DMI3*) genes was also increased, which helped to further regulate the *Armillaria* colonization in *G. elata* [1, 20]. Tan et al. conducted transcriptome analysis of the symbiotic region of *Armillaria* and *G. elata*, and found that the genes for a variety of extracellular enzymes (cellulase, xylanase, laccase) were in a state of down-regulation after *Armillaria* invaded the endothelial layer of *G. elata* [21]. At this time, the infestation of *Armillaria* on *G. elata* reached a state of equilibrium, there is no tendency to further invade the *G. elata*, providing a basis for understanding the symbiotic relationship between *Armillaria* and *G. elata* at the level of gene expression.

Currently, studies exploring the symbiotic relationship between *Armillaria* and *G. elata* have focused on phenotyping, physiological and biochemical assays, and transcriptome analyses. However, few studies have been reported to explore the symbiotic relationship between *Armillaria* and *G. elata* at the genome-wide level. Therefore, carrying out the whole genome analysis of the genus *Armillaria* is of theoretical significance and applied value for further understanding the role of *Armillaria* in symbiotic relationships with plants such as *G. elata* and *Polyporus umbellatus*. In this study, we present the draft genome of *A. gallica* M3 strain collected from Xiaocaoba Town, Yiliang County, Zhaotong City, Yunnan Province, and study the genome of *A. gallica* M3 with 12 other strains of Agaricales. This study helps to analyze the intrinsic mechanism that *A. gallica* M3 strain can be symbiotic with *G. elata* at the genomic level, and provides a theoretical basis for the standardization of *A. gallica* M3 strain accompanied by planting *G. elata*.

Materials and methods

Strains and culture conditions

Armillaria gallica M3 was provided by Yunnan Senhao Fungi Industry Co., Ltd. Routinely, *A. gallica* M3 was cultured in PDA solid medium (Huankai Microbial, Guangdong Province, China.) at 28 °C for 5 days in a dark environment, and reinoculated in fresh PDA solid medium at 28 °C. The strains used in this study are stored at the Faculty of Life Science and Technology, Kunming University of Science and Technology, Kunming city, China.

DNA extraction

Total genomic DNA was extracted using the CTAB hexadecyltrimethylammonium bromide method [22]. DNA purity was confirmed spectrophotometrically at 260 nm

with NanoDrop One spectrophotometer, and DNA concentration was determined with Qubit 3.0 fluorometer (Life Technologies, Carlsbad, CA, USA) to determine DNA concentration for accurate quantification of DNA (Thermo Fisher Scientific).

Whole-genome sequencing, assembly, and annotation

High-throughput sequencing of the genome was performed using the Illumina NovaSeq 6000 platform and the PromethION P48 sequencer (Oxford Nanopore Technologies, Oxford, UK). Preliminary genome assembly was first performed using the software NECAT (<https://github.com/xiaochuanle/NECAT>). The initial assembly was corrected by the software Racon (V1.4.11) and Pilon (V1.23) [23]. The error-corrected genome was de-hybridized using the software Purge haplotigs (V1.0.4) to obtain the final assembly (Draft Genome). The software Oxford Nanopore GUPPY (V4.0.2) was used to convert the raw image data files obtained from high-throughput sequencing into fast5 raw data by Base Calling [24]. After filtering low quality sequences, splice sequences, etc., the results were stored in FASTQ (fq) file format and annotated for analysis [25].

To evaluate the completeness of the assembled fungal genome and assess the quality of the gene set, we utilized BUSCO (Benchmarking Universal Single-Copy Orthologs) version 5.2.2. BUSCO provides a quantitative measure of genome and gene set completeness by identifying evolutionarily conserved single-copy orthologs. BUSCO evaluation is generally greater than 90%, and genome annotation is considered to be better. The lineage dataset is: fungi_odb10. BUSCO analysis first obtains data sets, and then provides quantitative measurement of genome and genome integrity by identifying evolutionarily conserved single-copy homologues.

Use RepeatMasker software (version: open-4.0.9) to make repeated comments based on RepBase library (<http://www.girinst.org/repbase>); Then use RepeatModeler software (version: open-1.0.11) to build a database for denovo prediction based on its own sequence characteristics, and also use RepeatMasker software (version: open-4.0.9) to compare and predict repeated sequences.

The tRNAscan-SE (V1.23) was used to find tRNA sequences in the genome based on the structural features of tRNAs [26]. The rRNA prediction was performed using the rRNA database. INFERNAL (V1.1.2) was used to find ncRNA sequences in the genome based on the Rfam database [27].

We use the following databases for gene function annotation: Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Pfam and NR. The protein sequences encoded by genes in gene set were compared with existing protein databases Uniprot, NR and

metabolic pathway database KEGG by diamond blastp (version: 2.0.11.149; Parameter:--evalue 1e-5) to obtain the functional information of the sequence and the metabolic pathway information that the protein may participate in. Among them, KEGG annotations are related to KEGG ORTHOLOGY and PATHWAY using Kobas (version: 3.0). Uniprot database records the corresponding relationship between each protein family and the functional nodes in Gene Ontology, and predicts the biological function of the protein sequence encoded by the gene through this system. Protein's conserved sequences, motif and domains were obtained by comparing the Pfam database with InterProScan (version: 5.52–86.0).

Pathogen Host Interactions (PHI) and Carbohydrate-active enzymes (CAZymes) database annotations

Pathogenicity analysis and annotation of target protein sequences using Diamond blastp based on the Pathogenic Host Interaction Database [28]. Using CAZy database to annotate the genome of *Armillaria gallica* M3 strain, so as to identify and classify CAZymes (<http://www.cazy.org/>).

Genome evolution analysis

We selected seven strains belonging to the genus *Armillaria* from NCBI database: *Armillaria gallica* Ar21-2, *Armillaria gallica* 012 m, *Armillaria borealis* ab13-tr4-ip16, *Armillaria ostroyae* C18/9, *Armillaria solidipes* 28–4, *Amanita brunnescens* Koide BX004, *Armillaria gallica* M3 and 6 fungal strains, and their protein sequences were obtained from NCBI. The specific information is shown in Table 1. Gene family clustering of 13 fungal strains using orthofinder software

Table 1 Fungal strains and Genebank registration numbers

Strains name	Strain number	Genebank No
<i>Armillaria gallica</i>	M3	
<i>Pleurotus eryngii</i>	ATCC 90797	GCA_015484515.1
<i>Armillaria gallica</i>	Ar21-2	GCA_002307695.1
<i>Armillaria gallica</i>	012 m	GCA_012064365.1
<i>Armillaria borealis</i>	AB13-TR4-IP16	GCA_013427175.2
<i>Armillaria ostoyae</i>	C18/9	GCA_900157425.1
<i>Armillaria solidipes</i>	28–4	GCA_002307675.1
<i>Amanita brunnescens</i>	Koide BX004	GCA_001691785.3
<i>Cylindrobasidium torrendii</i>	FP15055 ss-10	GCA_000934385.1
<i>Gymnopus luxurians</i>	FD-317 M1	GCA_000827265.1
<i>Coprinopsis cinerea</i>	AmutBmut pab1-1	GCA_016772295.1
<i>Coprinopsis marcescibilis</i>	CBS 121175	GCA_004369085.1
<i>Laccaria amethystina</i>	LaAM-08–1	GCA_000827195.1
<i>Laccaria bicolor</i>	S238N-H82	GCA_000143565.1

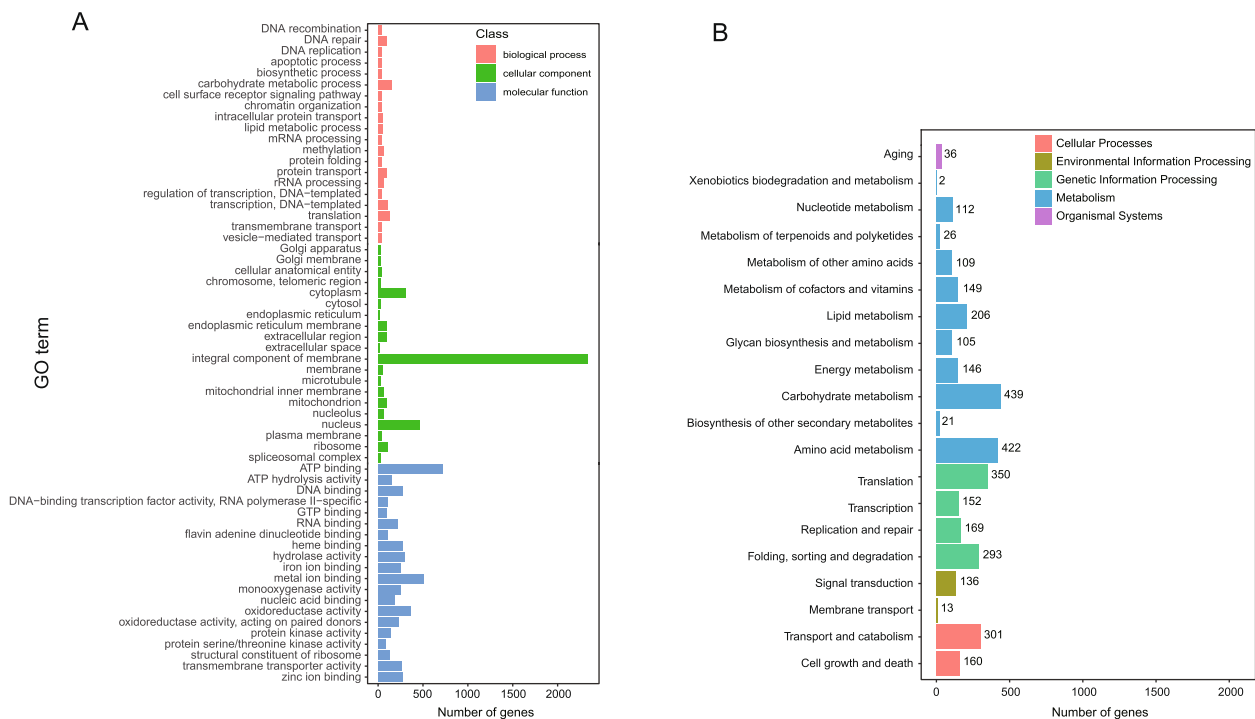


Fig. 1 GO functional annotation (A) and KEGG pathway annotation (B) of the *A. gallica* M3 genome

(V2.3.12) [29]. Among them, blastp (V2.6.0) software was used to do the comparison [30]. The software muscle (V3.8.31) was used to do multiple sequence alignment of protein sequences for each single-copy gene family. The results were filtered and concatenated into super gene using trimal (V1.4.rev22) [31, 32]. Finally, maximum likelihood (ML) species phylogenetic tree containing 1000 bootstrap was constructed using RAXML (V8.2.10) software (model PROTGAM-MAWAG) based on super gene [33].

Birth-mortality models in CAFE (V3.1) software was used to predict the contraction and expansion of the *A. gallica* gene family relative to its ancestors based on species evolutionary trees and gene family clustering results [34].

Comparative genomic analysis of *A. gallica* M3, *A. gallica* Ar21-2, *A. gallica* 012 m

Based on all amino acid sequences of the selected 3 species, gene families were clustered using Orthofinder software. Venn diagrams were drawn based on the statistics of the clustering results of the protein sequences of the three selected species. GO and KEGG analyses were performed using the R package clusterProfiler based on shared gene and uniq gene and functional annotation results, respectively [35–37].

Results

General genomic characteristics

The genome size of *A. gallica* M3 was 86,179,563 bp, with an average GC content of 47.43%. A total of 12,557 protein-coding sequences (CDSs) were identified. The genome completeness was 95.4% based on the BUSCO assessment, while 14 BUSCOs were fragmented and 21 BUSCOs were missing. In addition, 281 tRNAs, 60 rRNA manipulators (15 18S rRNAs, 28 5.8S rRNAs, 17 5S rRNAs) and 25 snRNAs (including 7 CD-boxes and 18 splicing) were predicted in the genome. The whole genome sequence of *A. gallica* M3 has been submitted to GenBank with the accession number is PRJNA1082085.

Gene annotation

In the GO analysis, a total of 7,671 genes were annotated in *A. gallica* M3, and were mapped to 18,679 GO terms. These include 4,058 biological process (BP), 5,048 cellular component (CC), and 9,573 molecular function (MF) (Fig. 1A). For the BP domain, carbohydrate metabolic process (153), translation (124), transcription, DNA-templated (109) are the most abundant terms. For the CC domain, genes were significantly enriched in integral component of membrane (2339), nucleus (467), and cytoplasm (308). For the MF domain, ATP binding (721), metal ion binding (505), and oxidoreductase activity (358) are the most abundant terms.

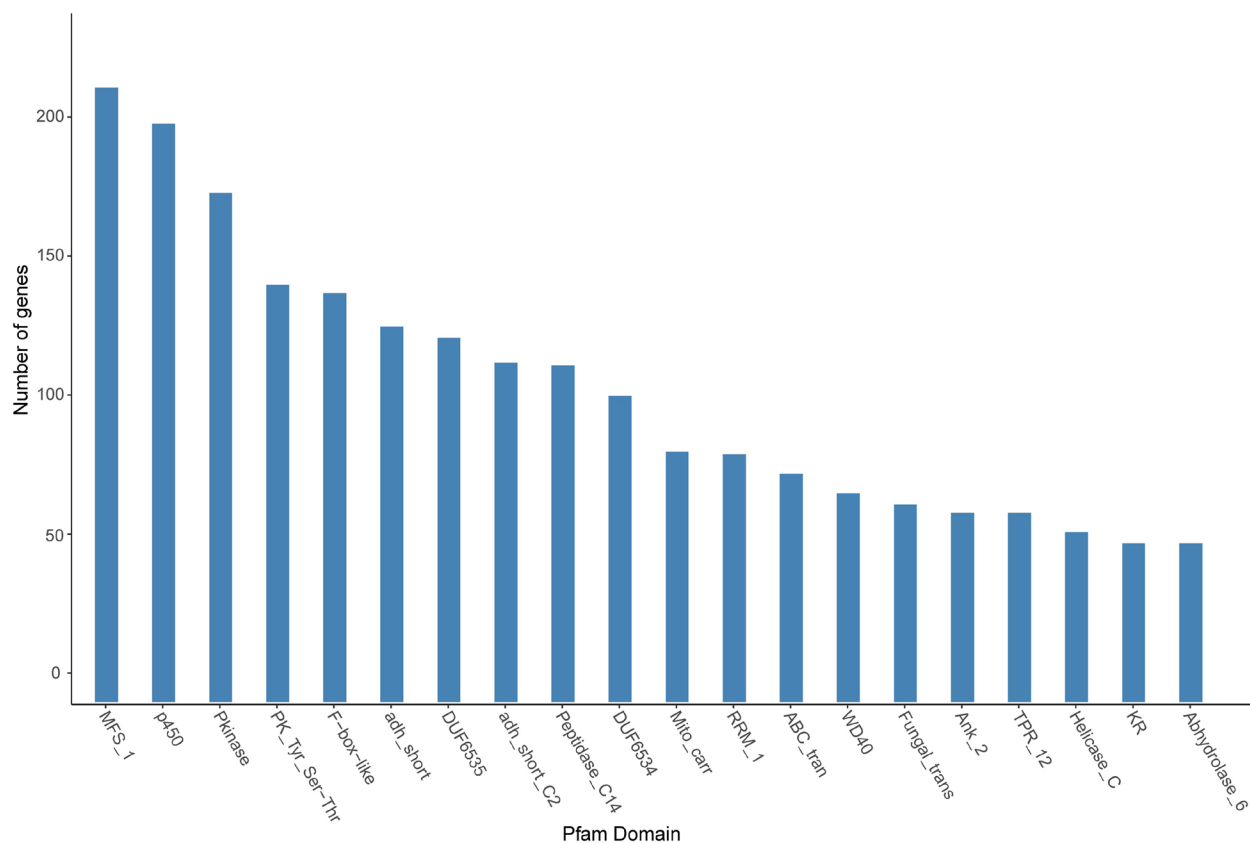


Fig. 2 Pfam functional annotation of the *A. gallica* M3 genome

In order to further reveal the biological functions of the coding genes in *A. gallica* M3, the genes were annotated in the KEGG database. 2,557 coding genes of *A. gallica* M3 were annotated to five major categories through the KEGG database. These include Cellular Processes, Environmental Information Processing, Genetic Information Processing, Metabolism, and Organic Systems. Among these categories, Metabolic pathways had the highest number of annotated genes (1737), with a high number of genes for Carbohydrate metabolism, Amino acid metabolism and Translation (Fig. 1B).

In the Pfam annotation, 7308 genes were annotated. Based on the annotation of Pfam domains, the genes annotated in each domain are statistically summarized, and the top 20 domains with the most gene annotations are shown in Fig. 2. Among them, these genes are significantly enriched in three protein families: Major Facilitator Superfamily, Cytochrome P450 and Protein kinase domain.

A total of 12,069 genes were annotated by searching the NR database. Among them, 62.68% of the genes were annotated to *Armillaria gallica*, 18.65% to *Armillaria ostoyae* and 16.38% to *Armillaria solidipes* (Fig. 3), and other species accounted for a small proportion.

The carbohydrate enzyme classification and annotation

A total of 452 CAZymes were annotated in the genome of *A. gallica* M3, which including 191 Glycoside hydrolases (GHs), 73 Glycosyltransferases (GTs), 22 Polysaccharide lyases (PLs), 119 Auxiliary Activities (AAs), 30 Carbohydrate Esterases (CEs) and 17 Carbohydrate-Binding Modules (CBMs) (Fig. 4A). Among them, GHs are a class of enzymes that hydrolyze glycosidic bonds. GTs are a class of enzymes that transport sugar molecules to other broad-spectrum receptors (Li et al., 2021).

According to the statistics, there are 47 GHs families in *A. gallica* M3, including 22 genes related to chitinase (Fig. 4B). To further reveal the biological functions of GHs in *A. gallica* M3, we annotated these genes in the KEGG database. 206 GHs of *A. gallica* M3 were involved in Metabolism in the KEGG database. Of these, 70 GHs were associated with Metabolic pathways, with the highest number of genes for Carbohydrate metabolism, Polysaccharide biosynthesis and metabolism and Lipid metabolism (Fig. 4C).

Pathogen Host Interaction (PHI)

A total of 3,412 genes were associated with PHI in *A. gallica* M3 by gene annotation. Among the 3,412 predicted

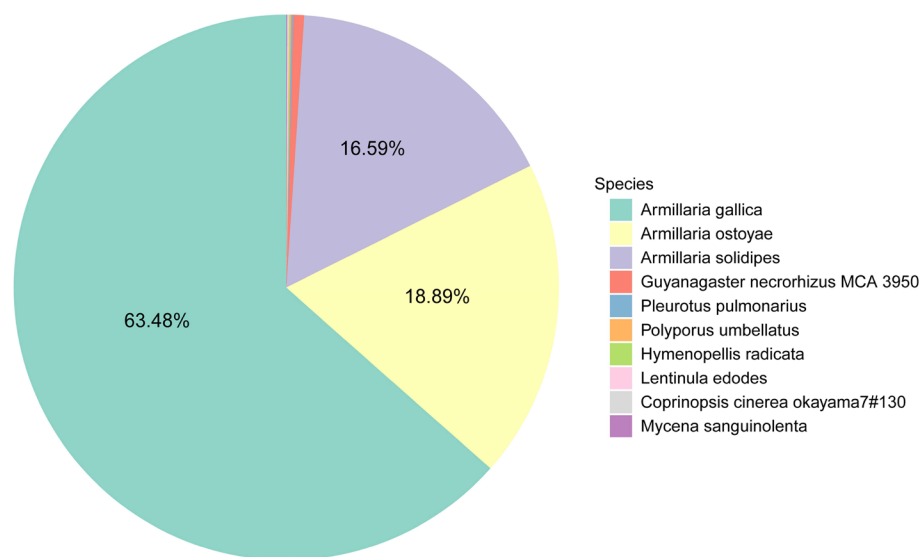


Fig. 3 NR functional annotation of the *A. gallica* M3 genome

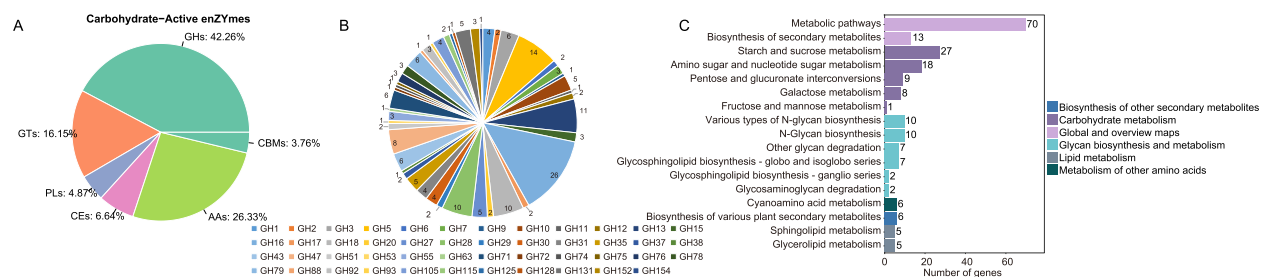


Fig. 4 Annotation on the CAZymes and GHs of *A. gallica* M3. **A** CAZymes classification chart. **B** GHs family classification chart. **C** Functional annotation chart of KEGG of GHs

genes, they were mainly distributed in Chemical target (0.35%), Reduced virulence (51.91%), Increased virulence (Hypervirulence) (4.63%), Enhanced antagonism (0.03%), Effector (plant avirulence determinant) (0.73%), Unaffected pathogenicity (31.33%), Loss of pathogenicity (6.95%) and Lethal (4.07%) in eight modules (Fig. 5).

Genome comparison analyses

A ML phylogenetic tree was constructed based on supergene using RAXML (version: 8.2.10) software (model PROTGAMMAWAG). The phylogenetic relationship between *A. gallica* M3 and other related species was analyzed. The phylogenetic tree showed that among the three different strains of *A. gallica*, *A. gallica* M3 was genetically distant from *A. gallica* Ar21-2 and was relatively closely related to *A. gallica* 012 m with 100% bootstrap value (Fig. 6).

Evolutionary analysis of gene families of *A. gallica* M3 was performed using CAFE (V3.1). The results of the expansion and contraction gene family analysis showed

that *A. gallica* M3 had 306 gene family expansions and 856 gene family contractions compared to the other 12 fungal strains. In the GO database, gene families such as oxidoreductase activity, fungal-type cell wall, monooxygenase activity, hydrolase activity, and extracellular region contracted significantly. Gene families such as transmembrane transporter protein activity and iron ion binding expanded significantly. In the KEGG database, gene families such as Glycine, serine and threonine metabolism and Tyrosine metabolism contracted significantly. Gene families such as Ubiquinone and other terpenoid – quinone biosynthesis, Biosynthesis of cofactors were significantly expanded (Fig. 7).

Analysis of the shared and unique gene family of *A. gallica* M3

This study counted and plotted Venn diagrams based on the clustering results of the protein sequences of *A. gallica* M3, *A. gallica* Ar21-2 and *A. gallica* 012 m (Fig. 8). The results showed that there were 7072 gene families in

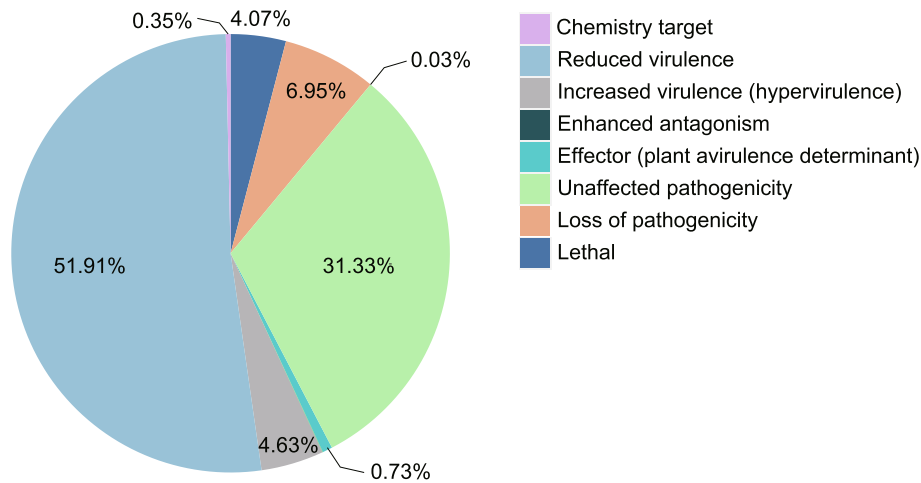


Fig. 5 Distribution of phenotypic categories of *A. gallica* M3 gene orthologs using the PHI database

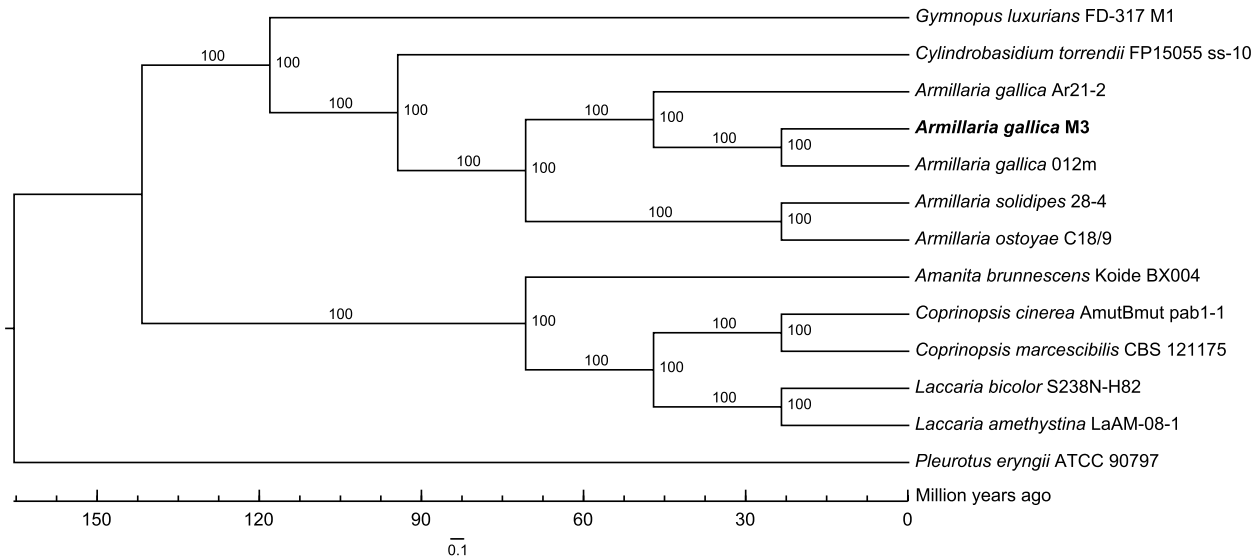


Fig. 6 Phylogenetic analysis of *A. gallica* M3

the three species, 719 gene families shared by *A. gallica* M3 with *A. gallica* Ar21-2, 624 gene families shared with *A. gallica* 012 m, and 412 gene families unique to *A. gallica* M3.

In order to further characterize *A. gallica* and *A. gallica* M3 strains themselves, GO and KEGG functional annotation analysis was performed on the shared genes of the three strains and the unique gene family of *A. gallica* M3. A total of 243 shared gene families and 90 unique gene families were coded and annotated in the GO analysis

(Fig. 9). Among them, the shared gene families such as carbohydrate metabolic process, flavin adenine dinucleotide binding were significantly enriched, and the unique gene families such as iron ion binding, monooxygenase activity were significantly enriched. In the KEGG analysis, 98 shared gene families and 70 unique gene families were encoded and annotated (Fig. 10). Among them, the shared gene families such as Biosynthesis of secondary metabolites, Biosynthesis of cofactors were significantly enriched, and the unique gene families such as Oxidative phosphorylation, Fatty acid biosynthesis were significantly enriched.

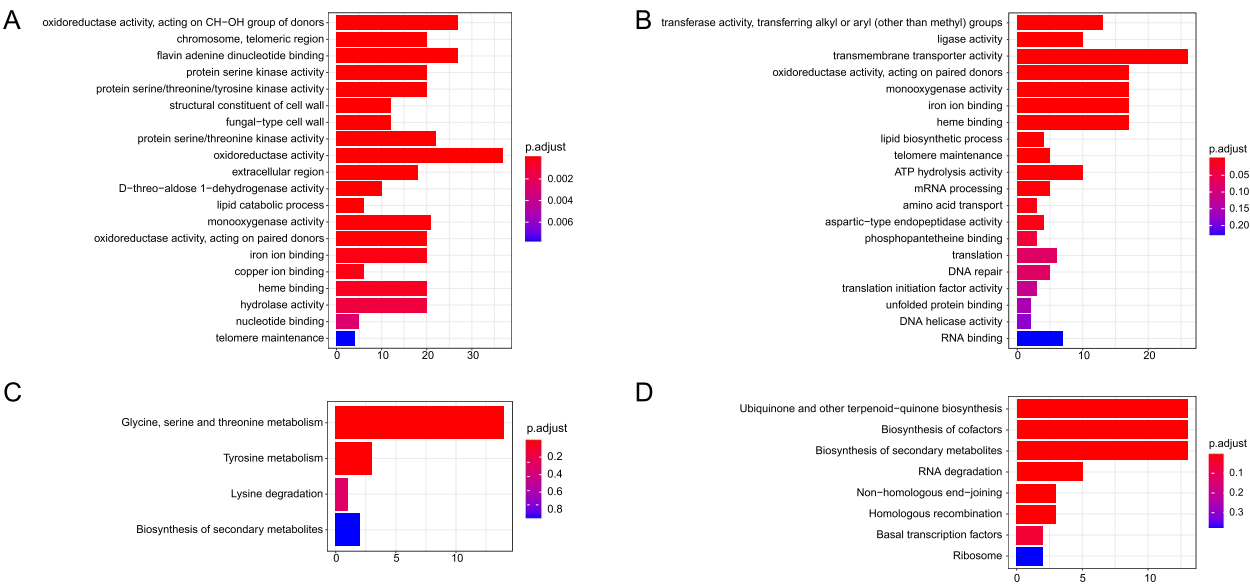


Fig. 7 Annotation of the gene family contraction and expansion of *A. gallica* M3. **A** contraction and **B** expansion gene family GO enrichment results. **C** contraction and **D** expansion gene family KEGG enrichment results

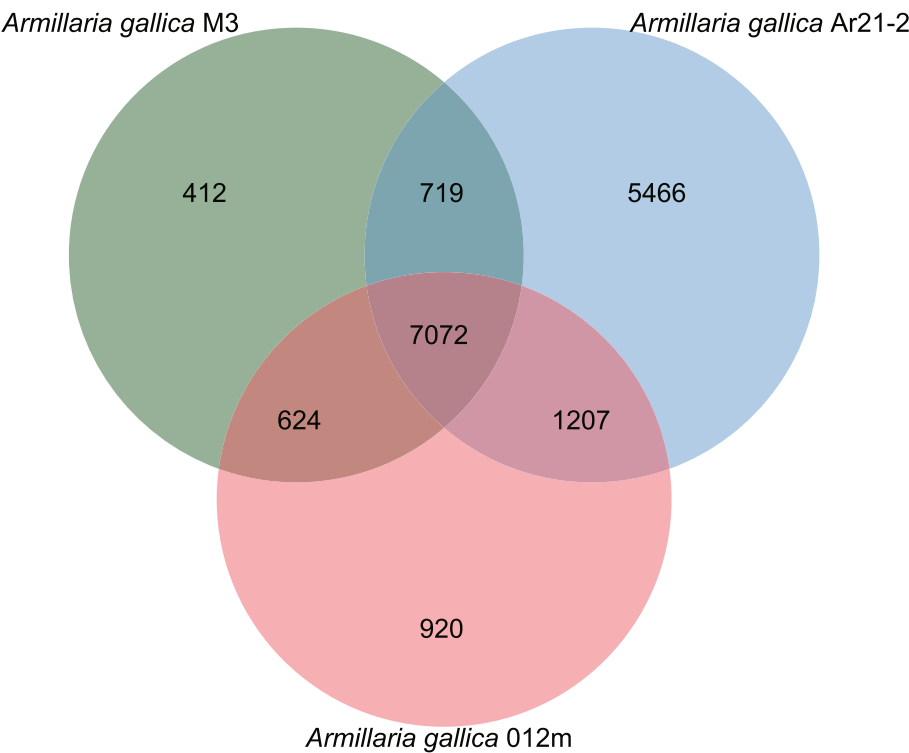


Fig. 8 Venn diagram of clustering result statistics of gene families of three species

Discussion

In this study, we obtained a 83.33 Mbp genome of *A. gallica* M3, which consisted of 38 contigs with an N50 of 6,065,498 bp and 47.43% GC content. A total of 12,557

genes were identified in the *A. gallica* M3 genome, and the Repetitive sequences accounted for about 44.36% of the genome. In the GO database, gene families related to monooxygenase activity associated with pathogenicity/

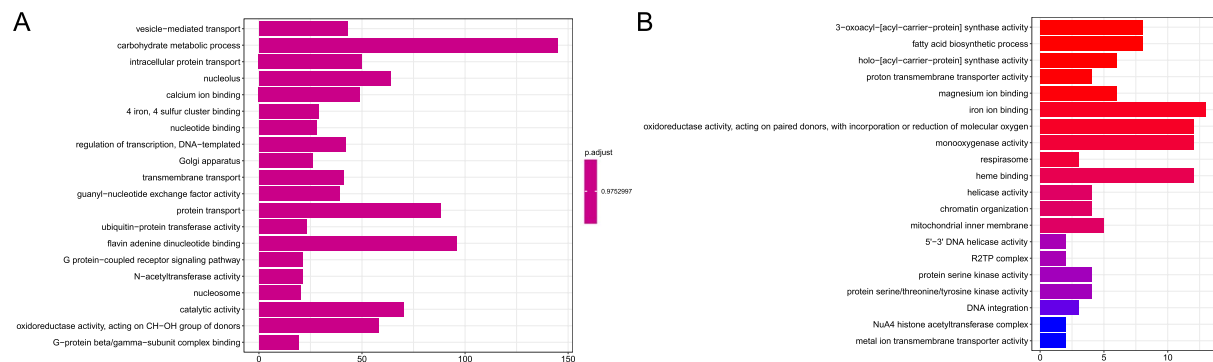


Fig. 9 GO enrichment results for shared (A) and unique (B) gene families

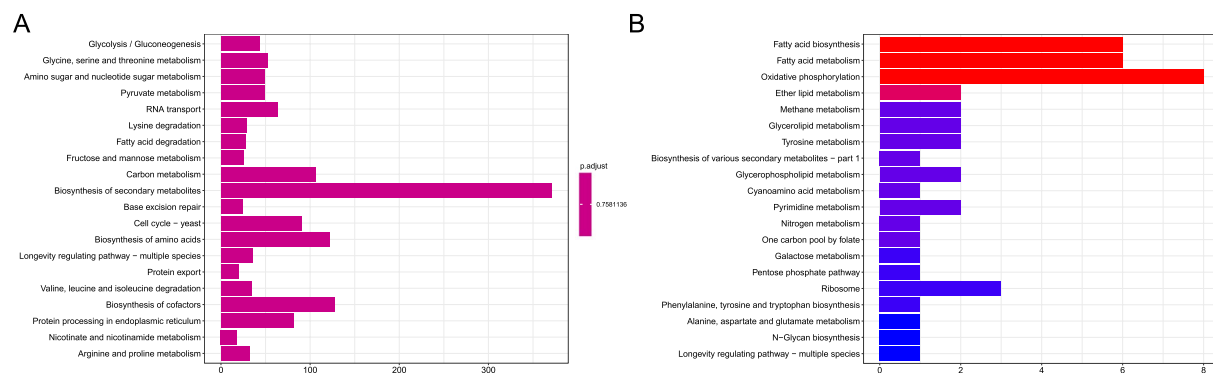


Fig. 10 KEGG enrichment results for shared (A) and unique (B) gene families

saprophytic, hydrolytic enzyme activity associated with infection ability and extracellular region were significantly contracted in the genome of *A. gallica* M3. Some studies have shown that the contraction of genes related to pathogenicity may enable the *Armillaria* to not further infect the endothelial layer of *G. elata* after infesting the epidermis of *G. elata*, which not only provides essential nutrients for the growth of *G. elata* but also avoids the pathological damage to *G. elata*, which is conducive to the symbiotic relationship between *Armillaria* and *G. elata* [38].

Carbohydrate-active enzymes (CAZymes) are crucial in the growth and development of plant pathogenic fungi [39]. The results predicted by the CAZy database indicate that there are 452 CAZymes in *A. gallica* M3 strain. *A. gallica* M3 had fewer CAZymes compared to the genome annotations of other *Armillaria* (Table 2). Among the 452 CAZymes, glycoside hydrolase and glycosyltransferase genes were significantly enriched, which are essential for the symbiosis of *Armillaria* with *G. elata*.

Glycoside hydrolases (GHs) are the main enzymes in the process of providing nutrients to *G. elata* or *Polyporus umbellatus* by *Armillaria*, which can hydrolyze glycosidic bonds and thus break down polysaccharides [40].

Playing a crucial role in the infestation of the epidermis of *G. elata* by the hyphae of *A. gallica* M3 strain, Collins et al. also showed that GHs may be associated with the pathogenicity of *Armillaria* [41]. Usually, plant symbiotic fungi contain more enzymes that degrade plant cell walls, such as Chitinase and Cellulase [42]. In *A. gallica* M3 strain containing 22 genes for Chitinase, it was reported that Chitinase, as a GHs, not only provides a material basis for the symbiosis between *A. gallica* and *G. elata*, but also senses the signal communication between *A. gallica* and *G. elata* [4]. This may facilitate the rhizomorphs to not cause further damage to the host after infestation of *G. elata*, so that the *A. gallica* and *G. elata* can better maintain the balance between infestation and nutrient acquisition in the process of symbiosis.

Glycosyltransferases (GTs) are enzymes that transport sugar molecules to other broad-spectrum receptors, where they complete the glycosylation of other molecules by creating glycosidic bonds [43]. Plants can produce polysaccharides by the action of GTs, including the synthesis and modification of cell wall polysaccharides, glycoproteins, glycolipids, and a number of small molecules, all catalyzed by GTs [44]. GTs within *Armillaria* have been reported to catalyze the transglycosylation of

Table 2 Number of genes encoding CAZymes in *Armillaria*

class	<i>A. gallica</i> M3	<i>A. gallica</i> 012m	<i>A. gallica</i> Ar21-2	<i>A. cepistipes</i> B5	<i>A. ostoyae</i> C18/9	<i>A. solidipes</i> 28-4
Total	452	824	826	810	764	789
GHs	191	331	318	329	300	302
GTs	73	100	106	97	92	93
PLs	22	25	25	27	31	29
CEs	30	142	144	136	135	150
CBMs	17	82	79	66	71	79
AAs	119	170	178	181	156	158

p-hydroxybenzyl alcohol compounds to produce gastrodin [45]. Tsai et al. also found that after *Armillaria* infected *G. elata*, the GTs involved in the gastrodin synthesis pathway appeared in the *Armillaria* infected *G. elata* [46]. This suggests that the *A. gallica* M3 provides nutrients to *G. elata* while its significantly enriched glycosyltransferase also promotes gastrodin synthesis.

There are nine families of Carbohydrate esterases (CEs) in *A. gallica* M3 strain, consisting of 30 genes (Table S1). Among them, the CE4 family has acetyl xylan, chitin, chitooligosaccharide and peptidoglycan deacetylase activities. It has been hypothesized that the chitin of pathogens invading plants can trigger recognition by plant defense and induce plant chitinase activity, resulting in the formation of lignin and healing tissue [47]. In addition, deacetylases in CE4 can modify plant chitinase, thereby rendering plant defenses ineffective, and modify chitin to chitosan, thereby enabling fungi to invade plants [41]. White-rot fungi are the only organisms capable of efficiently degrading lignin [48]. Among the carbohydrate active enzymes, GH1, GH3, GH5, GH6, and GH12 families are important enzymes involved in cellulose degradation. CE1, CE15, GH10, and GH43 are enzymes concerning xylan degradation. GH16 family proteins also have amylase and cellulase activities. *A. gallica* M3 also contains pectin cleavage family PL1 (pectin/pectin lyase), and PL3 (pectin lyase) genes.

Whole genome sequencing of *A. gallica* M3 strain contributes to the understanding of the function of relevant pathogenic genes of *A. gallica* during its interaction with *G. elata*. Previous studies have shown that *A. gallica* pathogenicity (lower virulence and aggressiveness) is weaker compared with *A. mellea*, *A. borealis* and *A. ostoyae*, a characteristic that may favor the establishment of a symbiotic relationship between *A. gallica* and *G. elata* rather than causing *G. elata* tuber disease [8, 49, 50]. A total of 3,412 PHI-associated genes were identified in the genome of *A. gallica* M3 by gene annotation, of which genes associated with Reduced virulence, Unaffected pathogenicity

and Loss of pathogenicity were significantly enriched. Among them, 51.91% of the genes were enriched to the Reduced virulence module, 6.95% of the genes were Loss of pathogenicity, while the genes related to the Lethal only accounted for 4.07%, which further indicated that *A. gallica* M3 strain was weak in pathogenicity and was suitable for mixing planting with *G. elata*. These results elucidate the characterization of *A. gallica* M3 strain and its ability to establish a symbiotic relationship with *G. elata* to some extent from a genomic point of view.

Conclusions

In this paper, we analyzed the whole genome information of *A. gallica* M3, and through gene annotation, we found that there were 3412 genes involved in the host pathogen interaction mechanism in *A. gallica* M3, 51.91% of the genes were enriched to the Reduced virulence module, and the genes related to the Lethal only accounted for 4.07%, which further indicated that *A. gallica* M3 strain was weak in pathogenicity and was suitable for mixing planting with *G. elata*. In addition, 42.26% of the genome was composed of glycoside hydrolases (GHs), 16.15% of the genome was composed of glycosyltransferases (GTs), which is helpful for *G. elata* to obtain the potential of nutritional application. In general, this study has important theoretical significance and application value for further understanding the diversity of *Armillaria*, determining the evolutionary history and phylogenetic relationship of *Armillaria*, and exploring the molecular mechanism of symbiosis between *Armillaria* and *G. elata*."

Acknowledgements

We are grateful for the Yunnan Major Scientific and Technological Projects. We acknowledge support by Yunnan Senhao Fungi Industry Co., Ltd.

Authors' contributions

L.C.Y. conducted experiments, data sorting and writing-preparation of original manuscript. L.Y. and S.L. conduct data analysis and review and edit manuscripts. L.J.J. and M.J.Y. assists in the analysis. L.Y.C. reviewed and revised the manuscript. L.L.B. designed the study, provided reagents, materials and revised papers. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Yunnan Major Scientific and Technological Projects (grant NO. 202202AG050008).

Data availability

The whole genome sequence of *A. gallica* M3 has been submitted to GenBank with the accession number is PRJNA1082085. The bioproject is PRJNA1082085. The BioSample is SAMN40199449. (<https://www.ncbi.nlm.nih.gov/bioproject/1082085>).

Declarations

Ethics approval and consent to participate

We collected *A. gallica* M3 from Yunnan Senhao Fungi Industry Co., Ltd. We collected *A. gallica* in accordance with relevant institutional, national and international guidelines and legislation.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 19 March 2024 Accepted: 14 October 2024

Published online: 31 March 2025

References

- Yuan Y, Jin X, Liu J, Zhao X, Zhou J, Wang X, Huang L. The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat Commun*. 2018;9(1):1615. Retrieved from <https://doi.org/10.1038/s41467-018-03423-5>.
- Wu Y, Wen S, Zhang W, Yu S, Yang K, Liu D, Sun J. *Gastrodia elata* Bl.: A Comprehensive Review of Its Traditional Use, Botany, Phytochemistry, Pharmacology, and Pharmacokinetics. *Evid Based Complement Altern Med*. 2023;2023:5606021. Retrieved from <https://doi.org/10.1155/2023/5606021>.
- Men J, Xing X, Guo S. A comparison of *Armillaria* spp. associated with *Polyporus umbellatus* and *Gastrodia elata* in China. *Mycosystema*. 2017;36(08):1072–82. Retrieved from <https://kns.cnki.net/kcms/detail/11.5180.Q.20170411.1048.007.html>. <https://doi.org/10.13346/j.mycosystema.170012>.
- Huang JW, Liu J, Zhou JH, Yuan Y, Yang Q, Huang LQ. Bioinformatics analysis and expression of chitinase genes in *Armillaria gallica*. *China J Chin Mater Med*. 2019;44(06):1151–5. Retrieved from <https://kns.cnki.net/kcms/detail/11.2272.R.20190129.2000.028.html>. <https://doi.org/10.19540/j.cnki.cjcmm.20190129.014>.
- Yu E, Gao Y, Li Y, Zang P, Zhao Y, He Z. An exploration of mechanism of high quality and yield of *Gastrodia elata* Bl. f. *glauca* by the isolation, identification and evaluation of *Armillaria*. *BMC Plant Biol*. 2022;22(1):621. Retrieved from <https://doi.org/10.1186/s12870-022-04007-8>.
- Xu JT. The changes of cell structure in the courses of *Armillaria mellea* penetrating the nutritional stems of *Gastrodia elata*. *Zhongguo yi xue ke xue yuan xue bao*. *Acta Academiae Medicinae Sinicae*. 2001;23(2):150–3.
- van der Heijden, MGA, Martin FM, Selosse MA, Sanders IR. Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytologist*. 2015;205(4):1406–1423. Retrieved from: <https://doi.org/10.1111/nph.13288>.
- Baumgartner K, Coetzee MPA, Hoffmeister D. Secrets of the subterranean pathosystem of *Armillaria*. *Mol Plant Pathol*. 2011;12(6):515–34. Retrieved from <https://doi.org/10.1111/j.1364-3703.2010.00693.x>.
- Smith ML, Bruhn JN, Anderson JB. The fungus *Armillaria bulbosa* is among the largest and oldest living organisms *Nature*. 1992;356(6368):428–31. Retrieved from <https://doi.org/10.1038/356428a0>.
- Sipos G, Anderson JB, Nagy LG. *Armillaria*. *Curr Biol*. 2018;28(7):R297–8. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0960982218300289>. <https://doi.org/10.1016/j.cub.2018.01.026>.
- Wang CH, Zeng YJ, Hou YB, Zeng CH, Wang M, Wang YH. Effects of *Armillaria* species on growth and quantity of active medicinal components of *G. elata* BL. f. *elata* tubers along an altitude gradient: Evidence from empirical experiments. *Phytochemistry Letters*. 2023;54:101–6. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1874390023000174>. <https://doi.org/10.1016/j.phytol.2023.01.016>.
- Liu T, Zhang W, Wang Z, Yang Y, Xu H, Song M, Bao H. Law of Content Change of Extracellular Enzymes and Polysaccharides in *Armillaria gallica*. *J Chinese Med Mater*. 2019;42(01):57–61. Retrieved from <https://kns.cnki.net/kcms/detail/44.1286.r.20190128.1058.012.html>. <https://doi.org/10.13863/j.jissn1001-4454.2019.01.012>.
- Zhao Q, Wu P, Liu H, Gui M. Progress in the study of symbiotic relationship between *Armillaria* and *Gastrodia elata*. *Yunnan Agric Sci Technol*. 2022;(02):56–58. https://kns.cnki.net/kcms2/article/abstract?v=vSxt4VHTcs5SHmK_ip0WqhN35oOmnAFBwHTeOpU4vmxP22l2y25bbey6ccILXhHcypqhG2rpfw6ZENOGSlo2rkONb7h2uy54dDtWti4hT-tP0q285yTLpJJ8r88b1BIHDWmVORZWCZw5rt4q3LaVjVd2jrAhJnSL_MF09Z-sdHFgLkmmZkuDeMQkmZHwwTVWSx-CAPtOs=&uniplatform=NZKPT&language=CHS.
- An S, Lu W, Zhang Y, Yuan Q, Wang D. Pharmacological Basis for Use of *Armillaria mellea* Polysaccharides in Alzheimer's Disease: Antia-poptosis and Antioxidation. *Oxid Med Cell Longev*. 2017;2017:4184562. Retrieved from <https://doi.org/10.1155/2017/4184562>.
- Sun H, Shu F, Guan Y, Kong F, Liu S, Liu Y, Li L. Study of anti-fatigue activity of polysaccharide from fruiting bodies of *Armillaria gallica*. *Int J Biol Macromolecules*. 2023;124611. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0141813023015052>. <https://doi.org/10.1016/j.ijbiomac.2023.124611>.
- Wu J, Zhou J, Lang Y, Yao L, Xu H, Shi H, Xu S. A polysaccharide from *Armillaria mellea* exhibits strong in vitro anticancer activity via apoptosis-involved mechanisms. *Int J Biol Macromol*. 2012;51(4):663–7. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0141813012002565>. <https://doi.org/10.1016/j.ijbiomac.2012.06.040>.
- Xu Q, Liu Y, Wang X, Gu H, Chen Z. Purification and characterization of a novel anti-fungal protein from *Gastrodia elata*. *Plant Physiol Biochem*. 1998;36(12):899–905. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0981942899800084>. [https://doi.org/10.1016/S0981-9428\(99\)80008-4](https://doi.org/10.1016/S0981-9428(99)80008-4).
- Cox KD, Layne DR, Scorza R, Schnabel G. *Gastrodia* anti-fungal protein from the orchid *Gastrodia elata* confers disease resistance to root pathogens in transgenic tobacco. *Planta*. 2006;224(6):1373–83. Retrieved from <https://doi.org/10.1007/s00425-006-0322-0>.
- Kretzschmar T, Kohlen W, Sasse J, Borghi L, Schlegel M, Bachelier JB, Martinoia E. A petunia ABC protein controls strigolactone-dependent symbiotic signalling and branching. *Nature*. 2012;483(7389):341–4. Retrieved from <https://doi.org/10.1038/nature10873>.
- Delaux PM, Xie X, Timme RE, Puech-Pages V, Dunand C, Lecompte E, Séjalon-Delmas N. Origin of strigolactones in the green lineage. *New Phytologist*. 2012;195(4):857–71. Retrieved from <https://doi.org/10.1111/j.1469-8137.2012.04209.x>.
- Tan Y, Bao Y, Cao J, Zhang H, Chen L, Xu H, Li K. Transcriptome analysis on symbiotic molecular mechanism of *Armillaria mellea* and *Gastrodia elata*. *Chin Tradit Herbal Drugs*. 2018;49(17):4125–30.
- Huang X, Duan N, Xu H, Xie TN, Xue YR, Liu CH. CTAB-PEG DNA Extraction from Fungi with High Contents of Polysaccharides. *Mol Biol (Mosk)*. 2018;52(4):718–26. <https://doi.org/10.1134/s0026898418040080>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Earl AM. Pilon an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e12963. <https://doi.org/10.1371/journal.pone.0112963>.
- Ewing B, Green P. Base-calling of automated sequencer traces using phred. II Error probabilities. *Genome Res*. 1998;8(3):186–94. <https://doi.org/10.1101/gr.8.3.186>.
- Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. 2010;38(6):1767–71. <https://doi.org/10.1093/nar/gkp1137>.

26. Chan PP, Lin BY, Mak AJ. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 2021;49(16):9077–96. <https://doi.org/10.1093/nar/gkab688>.
27. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 2013;29(22):2933–5. <https://doi.org/10.1093/bioinformatics/btt509>.
28. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12(1):59–60. <https://doi.org/10.1038/nmeth.3176>.
29. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
30. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421>.
31. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792–7. <https://doi.org/10.1093/nar/gkh340>.
32. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25(15):1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
33. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3. <https://doi.org/10.1093/bioinformatics/btu033>.
34. Han MV, Thomas GW, Lugo-Martínez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* 2013;30(8):1987–97. <https://doi.org/10.1093/molbev/mst100>.
35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM. Gene Ontology: tool for the unification of biology. *Nature Genet.* 2000;25(1):25–9. <https://doi.org/10.1038/75556>.
36. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
37. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Yu G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb).* 2021;2(3):100141. <https://doi.org/10.1016/j.xinn.2021.100141>.
38. Zhan M, Tian M, Wang W, Li G, Lu X, Cai G, Huang L. Draft genomic sequence of *Armillaria gallica* 012m: insights into its symbiotic relationship with *Gastrodia elata*. *Bras J Microbiol.* 2020;51(4):1539–52. Retrieved from <https://doi.org/10.1007/s42770-020-00317-x>.
39. Hage H, Rosso MN. Evolution of Fungal Carbohydrate-Active Enzyme Portfolios and Adaptation to Plant Cell-Wall Polymers. *J Fungi.* 2021;7(3). <https://doi.org/10.3390/jof7030185>.
40. Li B, Liu L, Shan TT, Guo SX. Analyses of the glycoside hydrolase family genes and their differential expressions between rhizomorphs and hypha in *Armillaria mellea*. *Mycosystema.* 2021;40(06):1369–1379+1373–1383. Retrieved from <https://kns.cnki.net/kcms/detail/11.5180.Q.20210603.1550.001.html>. <https://doi.org/10.13346/j.mycosystema.210078>.
41. Collins C, Keane TM, Turner DJ, O'Keeffe G, Fitzpatrick DA, Doyle S. Genomic and proteomic dissection of the ubiquitous plant pathogen, *Armillaria mellea*: toward a new infection model system. *J Proteome Res.* 2013;12(6):2552–70. <https://doi.org/10.1021/pr301131t>.
42. Fradin EF, Thomma BP. Physiology and molecular aspects of Verticillium wilt diseases caused by *V. dahliae* and *V. albo-atrum*. *Mol Plant Pathol.* 2006;7(2):71–86. <https://doi.org/10.1111/j.1364-3703.2006.00323.x>.
43. Li CW, Zhang H, Rao P, Zheng DR, Wang Y, Li YH. Research progress on biosynthesis pathway of pentacyclic triterpenoids in plants. *Chin Tradit Herbal Drugs.* 2021;52(11):3436–52. Retrieved from <https://kns.cnki.net/kcms/detail/12.1108.R.20210315.1534.004.html>.
44. Zhou PN Li Y, Pu TZ, Yu K, Zhang XQ, Gong L. Cloning and bioinformatics analysis of β -1, 4-xylosyltransferase IRX9 gene from *Dendrobium huoshanense*. *Chinese Traditional and Herbal Drugs.* 2019;50(05):1212–9.
45. Liu XJ, Dong YC, Feng Y, Zhang HF, He GQ, Chen QH. Preliminary Study on the Identification of Glucosyltransferase from Biotransformation Broth of Gastrodin Synthesis by Cultured Cells of *Armillaria Luteo-Virens* Sacc. *J Chin Inst Food Sci Technol.* 2012;12(10):120–4. <https://doi.org/10.16429/j.1009-7848.2012.10.026>.
46. Tsai CC, Wu KM, Chiang TY, Huang CY, Chou CH, Li SJ, Chiang YC. Comparative transcriptome analysis of *Gastrodia elata* (Orchidaceae) in response to fungus symbiosis to identify gastrodin biosynthesis-related genes. *BMC Genomics.* 2016;17:212. <https://doi.org/10.1186/s12864-016-2508-6>.
47. Tsigos I, Bouriots V. Purification and characterization of chitin deacetylase from *Colletotrichum lindemuthianum*. *J Biol Chem.* 1995;270(44):26286–91. <https://doi.org/10.1074/jbc.270.44.26286>.
48. Levasseur A, Lomascolo A, Chabrol O, Ruiz-Dueñas FJ, Boukhris-Uzan E, Piumi F, Record E. The genome of the white-rot fungus *Pycnoporus cinnabarinus*: a basidiomycete model with a versatile arsenal for lignocellulosic biomass breakdown. *BMC Genomics.* 2014;15:486. <https://doi.org/10.1186/1471-2164-15-486>.
49. Morrison DJ. Rhizomorph growth habit, saprophytic ability and virulence of 15 *Armillaria* species. *Forest Pathol.* 2004;34(1):15–26. Retrieved from <https://doi.org/10.1046/j.1439-0329.2003.00345.x>.
50. Ota Y, Matsushita N, Nagasawa E, Terashita T, Fukuda K, Suzuki K. Biological Species of *Armillaria* in Japan. *Plant Dis.* 1998;82(5):537–43. Retrieved from <https://doi.org/10.1094/PDIS.1998.82.5.537>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.