



# Genome Sequencing and Analysis of the Fungal Symbiont of *Sirex noctilio*, *Amylostereum areolatum*: Revealing the Biology of Fungus-Insect Mutualism

Ningning Fu,<sup>a,b</sup> Ming Wang,<sup>a,b</sup> Lixiang Wang,<sup>a,b\*</sup> Youqing Luo,<sup>a,b</sup> Lili Ren<sup>a,b</sup>

<sup>a</sup>Beijing Key Laboratory for Forest Pest Control, Beijing Forestry University, Beijing, China

<sup>b</sup>Sino-French Joint Laboratory for Invasive Forest Pests in Eurasia, Beijing Forestry University—INRAE, Beijing, China

Youqing Luo and Lili Ren contributed equally to this article. Author order was determined on the basis of seniority.

**ABSTRACT** *Amylostereum areolatum* is the symbiotic fungus of the Eurasian woodwasp, *Sirex noctilio*, a globally invasive species. The mutualistic symbiont is associated with the woodwasp, assisting the damage process and providing nutrition for its insect partners. Colonization and growth of *A. areolatum* have essential impacts on the development and spread of *S. noctilio*, though the mechanism of interaction between the two has been poorly described. In this study, the first genome of this symbiotic fungus was sequenced, assembled, and annotated. The assembled *A. areolatum* genome was 57.5 Mb (54.51% GC content) with 15,611 protein-coding genes. We identified 580 carbohydrate-active enzymes (CAZymes), 661 genes associated with pathogen-host interactions, and 318 genes encoding transport proteins in total. The genome annotation revealed 10 terpene/phytoene synthases responsible for terpenoid biosynthesis, which could be classified into three clades. Terpene synthase gene clusters in clade II were conserved well across Russulales. In this cluster, genes encoding mevalonate kinase (MK), EGR12 (COG1557), and nonplant terpene cyclases (cd00687) were the known biosynthesis and regulatory genes. Genome sequence analysis of this fungus would prove the possibility of *A. areolatum* volatiles affecting the host selection of *S. noctilio* on a molecular basis. We further clarified that *A. areolatum* was a strict obligate symbiotic fungus. The wasps might protect the fungus before it was introduced into a suitable host substrate by oviposition, while the fungus would provide *S. noctilio* with a suitable environment and nutrients for the larval growth. These results would lay a foundation for our understanding of the mechanism of this entomogenous symbiosis.

**IMPORTANCE** *Sirex noctilio* (F.), together with *Amylostereum areolatum*, a wood-decaying symbiotic fungus, causes severe damage to *Pinus* species worldwide. In China, it causes extensive death of Mongolian pine (*Pinus sylvestris* var. *mongolica*). There is an obligate dependency mutualism between the woodwasp and its fungus. Studies have suggested that the fungal growth rate affected the size of the wasps: larger adults emerged from sites with a higher fungus growth rate. This genome is the first reported genome sequence of a woodwasp symbiotic fungus. Genome sequence analysis of this fungus would prove the possibility of *A. areolatum* volatiles affecting the host selection of *S. noctilio* on a molecular basis. We further clarified that *A. areolatum* was a strict obligate symbiotic fungus and that it would provide *S. noctilio* with a suitable environment and with nutrients for the larval growth. These results would lay a foundation for our understanding of the mechanism of this entomogenous symbiosis.

**KEYWORDS** secondary metabolism, *Amylostereum areolatum*, *Sirex noctilio* (F.), cell wall-degrading enzymes, genome, virulence

**Citation** Fu N, Wang M, Wang L, Luo Y, Ren L. 2020. Genome sequencing and analysis of the fungal symbiont of *Sirex noctilio*, *Amylostereum areolatum*: revealing the biology of fungus-insect mutualism. mSphere 5:e00301-20. <https://doi.org/10.1128/mSphere.00301-20>.

**Editor** Garret Suen, University of Wisconsin—Madison

**Copyright** © 2020 Fu et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Youqing Luo, [youqingluo@126.com](mailto:youqingluo@126.com), or Lili Ren, [lily\\_ren@bjfu.edu.cn](mailto:lily_ren@bjfu.edu.cn).

\* Present address: Lixiang Wang, College of Plant Protection, Gansu Agricultural University, Lanzhou, China.

**Received** 1 April 2020

**Accepted** 28 April 2020

**Published** 13 May 2020

The Eurasian woodwasp, *Sirex noctilio* (F.) (Hymenoptera; Symphyta; Siricidae), is an exotic invasive pest that damages *Pinus* species. It is recognized as a high-risk invasive species by the North American Plant Protection Organization (NAPPO) and the United States Department of Agriculture (USDA) (1, 2). *S. noctilio* is native to Europe and North Africa, where woodwasp is generally considered a secondary pest and of negligible economic concern. However, it had spread and established in several countries in the Southern Hemisphere during the twentieth century and, later, in North America and Southern Africa (3–6). In China, it was first found in Daqing, Heilongjiang Province, in 2013, and has been spotted in many cities since then, causing substantial deaths in Mongolian pine (*Pinus sylvestris* var. *mongolica*) plantations (7).

*Amylostereum areolatum* (Fr.) Boidin (Basidiomycotina: Corticiaceae) is a fungal symbiont of *S. noctilio*. There is a very strict obligate dependency mutualism between woodwasp and its fungal symbiont *A. areolatum* (7). *S. noctilio* has a specialized organ for storing symbiotic fungus, the mycangium, that is a reliable indicator of reciprocal, mutually beneficial adaptations and, in many cases, of obligate dependencies (2). First, the mycangium protects the fungus before it is introduced into a suitable host substrate. Next, during oviposition, the female woodwasp inoculates its symbiotic fungus and phytotoxic venom mucus into the host tree with the egg through the ovipositor. Sometimes, when the host is not suitable, no egg is deposited and only the fungus and venom are (8–10). Starting from the second instar, the fungus migrates from the previous larval instar to the next. When the adult emerges, the fungus is taken up from the wall of the pupal chamber into its mycangium (7). In this way, the female collects the oidia of *A. areolatum* produced in the insect galleries for dissemination and inoculation into new trees. In turn, the fungus is also essential to the development and even the reproductive potential of *S. noctilio* (11–14). *A. areolatum* infects and dries the wood substrate to provide a more suitable microenvironment for egg and larvae. In addition, degradation of cellulose, hemicellulose, and pectin by rich repertoire enzymes of the symbiotic fungus is presumably crucial for the woodwasp larvae. The development of the woodwasp, and thus its reproductive potential, are tied to the vigor of its symbiotic fungus (2). Studies have shown that woodwasp larvae feed on *A. areolatum* until the third or fourth instar and then feed on white-rotted wood (14–16). Madden and Coutts (1979) suggested that the fungus growth rate affected the size of the woodwasps; larger adults emerged from sites with higher rates of fungus growth (12). Obviously, the woodwasp derives considerable benefits from its partnership with *A. areolatum*, and the benefits are extensively linked to the colonization and growth of the fungi in pine hosts. Carbohydrate-active enzymes (CAZymes) and virulence-related genes are essential in the colonization and growth of fungus. White rot fungi can efficiently degrade lignocellulosic biomass, especially that derived from plants, for their diverse CAZymes (17). Many virulence-related genes that have been studied were found to be associated with the manipulation of plant defenses to promote fungus infections (18). These genes play an important role in defense mechanisms, signal transduction, carbohydrate transport and metabolism, intracellular trafficking, secretion, and vesicular transport (18–20).

Volatile components play an important role in the attraction of female woodwasps to the plant host. A Y-tube olfactometer was used to assess the behavior preferences of adult female wasps with respect to volatiles (21). Previous research showed that *S. noctilio* showed a stronger positive response to volatile components of *A. areolatum* than to those of attacked pines. These fungal volatiles could also attract woodwasps to hosts infected by *A. areolatum* (21). Our laboratory analyzed the volatile organic compounds (VOCs) of *A. areolatum* and other endophytic fungi in Mongolian pine and found that sesquiterpenes attract mated *S. noctilio* females but not unmated ones (15). It may be more important for a mated female to locate a weakened tree that has been inoculated with the fungi needed for the development of the woodwasp (22, 23). Although terpenoid compounds may play an important role in spatial aggregation of woodwasps, no terpene biosynthetic enzyme has so far been described in *A. areolatum*. Genome sequencing methods are commonly used to discover biosynthetic pathways

of terpenoid compounds by identifying the genes involved in the activity of secondary metabolites. For example, the potential terpene synthase (TPS) genes in the wood-rotting fungi *Stereum hirsutum* and *Coprinus cinereus* were identified by genome sequencing, and their biochemical activities were subsequently characterized (24, 25). Similarly, the availability of genome sequences may lead to the discovery of the TPS gene family in *A. areolatum*.

Wood-feeding insects are facing many difficulties in obtaining nutrition from their host's wood resources. The European woodwasp is associated with its mutualistic fungus, *A. areolatum*, which assists host selection and provides nutrition for its insect partner, though the interaction mechanism for this symbiosis has been poorly described. In this study, we completed the genome sequencing of the symbiotic fungus of *S. noctilio* and analyzed carbohydrate enzymes, virulence genes, and secondary metabolism genes. These findings would help us to explain the hypothesis that there are enzymes secreted by *A. areolatum* playing crucial roles in supporting the life cycle of *S. noctilio*, such as by degrading plant cell walls to provide nutrients for the colonization and growth of *S. noctilio* larvae and by producing sesquiterpene compounds to attract the female to oviposit. Our results lay a foundation for a better understanding of the mechanism of mutualism between *S. noctilio* and *A. areolatum*.

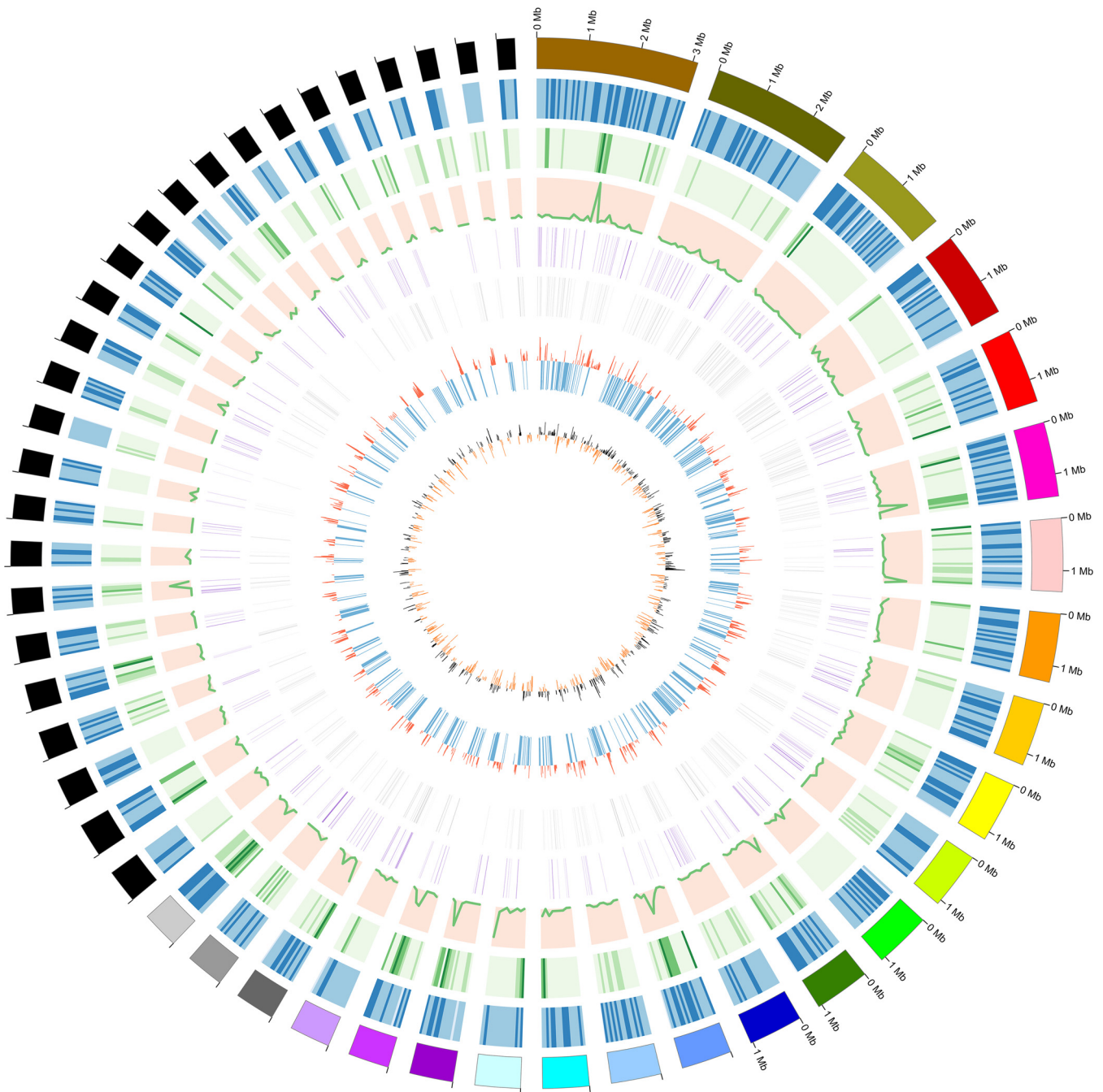
## RESULTS

**General genomic characteristics.** The *A. areolatum* genome was sequenced using a combination of Illumina and PacBio Sequel technologies. After quality control, we obtained 6,839 Mb of HiSeq data (118× coverage) and 4,314 Mb of PacBio data (74× coverage). Combined sequences from the two platforms were assembled into 248 scaffolds with an  $N_{50}$  value of 789 kb to obtain a total genome size of 53,481,184 bp (54.51% GC content) (Fig. 1) (Table 1). In addition, we predicted 15,611 genes, with an average length of 2,275 bp (Table 1). Among the protein-coding genes, 13,378 (85.69%) had significant sequence similarity to previously documented fungal sequences, whereas 14.31% of the genes did not have significant matches to any known genes. BUSCO was used to calculate the completeness of assembly and annotation. Among 290 single-copy orthologs, 92.8% of contigs were complete (232 complete single-copy BUSCOs and 37 complete duplicated BUSCOs), while only 1.4% were fragmented and 5.8% were missing.

In further analyses, we detected 97,258 exons with an average length of 245.35 bp (total length, 35,517,529 bp). The average length of the introns was 142.76 nucleotides. For noncoding RNAs (ncRNAs), 457 tRNA genes in the *A. areolatum* genome were identified using tRNAscan-SE (11). Among the tRNAs, 429 anticodon tRNAs corresponded to the 20 common amino acids and the other tRNAs represented possible pseudogenes.

**Repetitive elements and transposases.** Besides protein-coding gene sequences, an important portion of the fungal genome is the repetitive element. In the *A. areolatum* genome, we identified 14,513,406 bp (25.21% of the genomic sequence) of repetitive elements (Fig. 1; see also Table S1 in the supplemental material). Tandem repeat sequences accounted for 1.1% and transposable elements (TEs) for 24.11% of the assembled genome. Among the TEs, class I element (retrotransposon) TEs and class II element (DNA) TEs accounted for 15.50% and 1.36% of the genome, respectively. Unknown and other repetitive elements covered 8.17%. Long terminal repeats (LTRs) were the most abundant class I TEs (14.79% of the genomic sequence) and included the Copia, Gypsy, and BEL/Pao retrotransposon superfamilies.

**Phylogeny.** OrthoFinder identified 56,374 gene clusters, among which were included 913 orthologous genes shared among all the fungal species. From these shared gene clusters, 41 single-copy orthologous genes were chosen to analyze the evolutionary relationship of *A. areolatum* and other 82 Agaricomycetes reference genomes (Table S2). The maximum likelihood (ML) analysis identified, with high (>70%) bootstrap support, 15 major fungal clades of the orders Cantharellales, Auriculariales, Agaricales, Boletales, Amylocorticiales, Atheliales, Gloeophyllales, Jaapiales, Corticiales,



**FIG 1** Circular representation of genomic features of *A. areolatum*. The graph shows the genomic features of *Amylostereum areolatum* (Fr.) Boidin. The circles illustrate, from outside to inside, the scaffolds (sorted by length) of *A. areolatum*. The second and third circles show gene densities (gene number in 50,000-bp nonoverlapping windows) and repetitive sequences (repeat coverage length in 50,000-bp nonoverlapping windows). The fourth circle shows ncRNA densities (ncRNA number in 100,000-bp nonoverlapping windows). The fifth circle shows genes coding for CAZY proteins. The sixth circle contains members of the PHI gene group. The seventh and last circle indicates percent G+C content and GC skew (in 20,000-bp nonoverlapping windows).

Russulales, Polyporales, Hymenochaetales, Trechisporales, Geastrales, and Sebaciniales (Fig. 2). Phylogenetic analysis revealed that *A. areolatum* clustered with other Russulales species and close to the plant-pathogenic fungus *Peniophora* sp. strain CONT.

**Carbohydrate-active enzymes.** In the *A. areolatum* genome, we identified 580 CAZymes, consisting of 219 glycoside hydrolases (GHs), 67 glycosyl transferases (GTs), 16 polysaccharide lyases (PL), 60 carbohydrate esterases (CEs), 58 carbohydrate-binding modules (CBMs), and 160 enzymes with auxiliary activities (AAs) (Table S3). While the

**TABLE 1** Whole-genome assembly features of *Amylostereum areolatum* (Fr.) Boidin

Parameter or annotation	Value
Assembly parameters	
Total genome size (bp)	57,567,124
No. of scaffolds	248
Maximum scaffold length (bp)	3,099,494
Minimum scaffold length (bp)	365
Depth of genome coverage (×)	118
Scaffold $N_{50}$ length (bp)	731,813
GC content (%)	54.51
Gene annotations	
Total no. of genes	15,611
Total no. of annotated genes	13,378
Total no. of ncRNAs	722

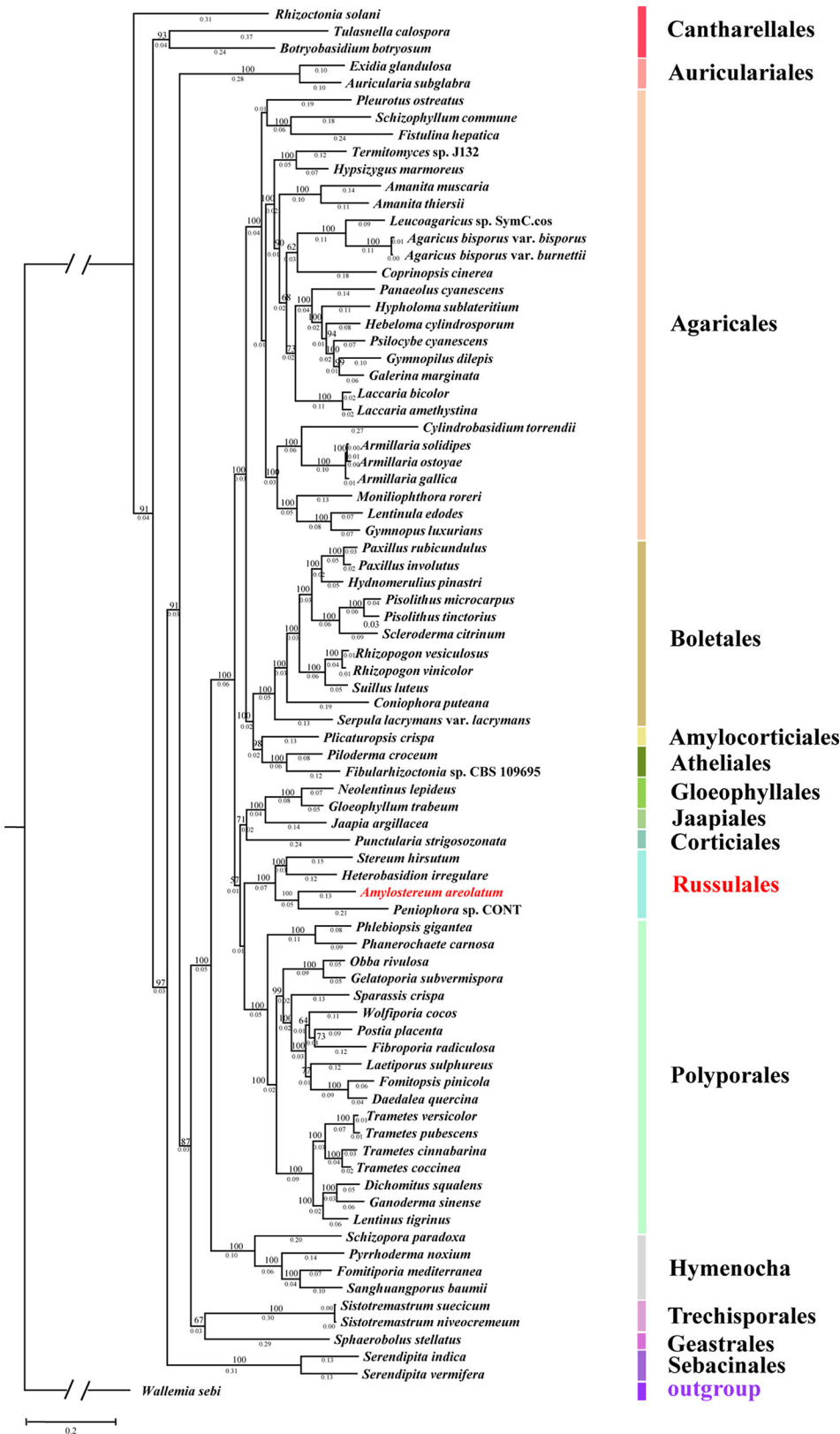
number of CAZymes in *A. areolatum* (580) was comparable to the number in other Russulales fungi, the number of CAZyme genes was comparable to the number in the forest pathogens *Heterobasidion parviporum* (519), *S. hirsutum* (543), *Heterobasidion irregulare* (442), and *Peniophora* sp. strain CONT (639). *A. areolatum* had more CAZyme genes than the traditional saprophytic mushrooms *Lactarius echinatus* (83), *Russula foetens* (90), *Lactarius deliciosus* (81), and *Albatrellus ellisii* (77) (Fig. 3).

There were more genes encoding GHs and AAs than other protein types in the *A. areolatum* genome. We found that the members of 45 GH gene families, including GH3 (15 copies), GH5 (18 copies), and GH18 (16 copies), were particularly abundant. In addition, we found two GH6 and eight GH7 members related to the degradation of crystalline cellulose. A classification of the AA family revealed that a majority of the AAs were members of the AA3 family (62 copies), including the subfamilies of cellobiose dehydrogenase (EC 1.1.99.18), glucose 1 oxidase (EC 1.1.3.4), aryl alcohol oxidase (EC 1.1.3.7), alcohol oxidase (EC 1.1.3.13), and pyranose oxidase (EC 1.1.3.10) (Fig. 1) (Table S3).

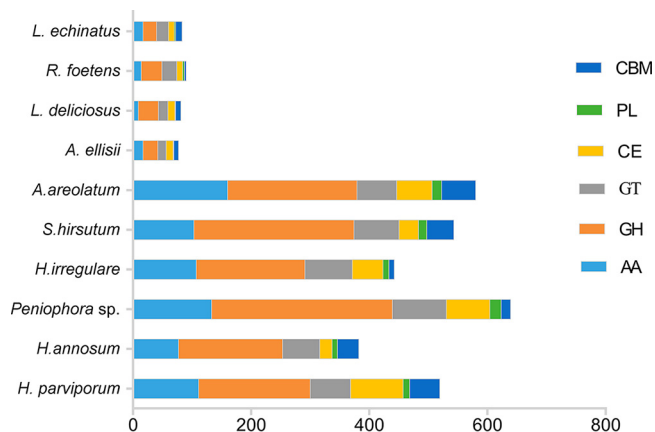
**Putative virulence-associated genes.** We found a total of 661 predicted *A. areolatum* genes associated with pathogen-host interactions (PHI). The highest proportion (45.39%) was related to “reduced virulence,” followed by “unaffected pathogenicity” (23.45%), “loss of pathogenicity” (10.14%), “lethal” (9.23%), “mixed outcome” (8.47%), “chemistry target” (1.21%), “increased virulence (hypervirulence)” (1.66%), and “effector (plant avirulence determinant)” (0.45%) (Fig. 1; see also Fig. 4) (Table S4). Besides, we annotated 47 genes that have been reported as virulence factors (VFs) in pathogenic bacteria (virulence factors of pathogenic bacteria [Virulence Factor Database {VFDB}]) (Table S5). These genes may be putative pathogenicity factors in *A. areolatum* based on their established role in pathogenesis in other host species. Furthermore, we identified 318 transport proteins belonging to 83 families in *A. areolatum*. Among these transporters, 37 genes encoded major facilitator superfamily (MFS) proteins and 18 encoded ATP-binding cassette (ABC) proteins; fewer genes encoded proteins in other families (see Fig. S2 in the supplemental material; see also Table S6).

**Sesquiterpene synthase clusters.** AntiSMASH analysis revealed that the number and type of secondary metabolite genes predicted for *A. areolatum* were comparable with those reported for other Russulales species, which carried 19 secondary metabolism gene clusters, including clusters of genes encoding eight terpene/phytoene synthases, two nonribosomal peptide synthetases (NRPS), one type-I polyketide synthase (T1PKS), and one siderophore synthase and seven unknown gene clusters (Table S7). It is worth mentioning that we identified the largest number of terpene/phytoene synthase gene clusters in the genome of *A. areolatum*. These enzymes might be involved in the biosynthesis of a terpenoid, although further investigation is needed to ascertain this.

To elucidate the conserved and diverged structures of the terpene biosynthetic clusters, we obtained two ortholog groups containing five sesquiterpene synthase (STS)



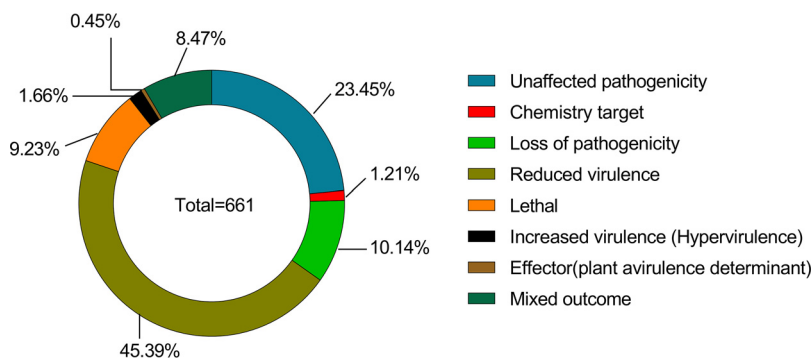
**FIG 2** Phylogenetic location of *A. areolatum* among Agaricomycetes. The topology of the phylogenetic tree was constructed by the maximum likelihood method (bootstrap = 1,000, LG+I+G+F model); *A. areolatum* is shown in red. Bootstrapping-based branch support and a scale bar representing the mean number of amino acid substitutions per site are shown.



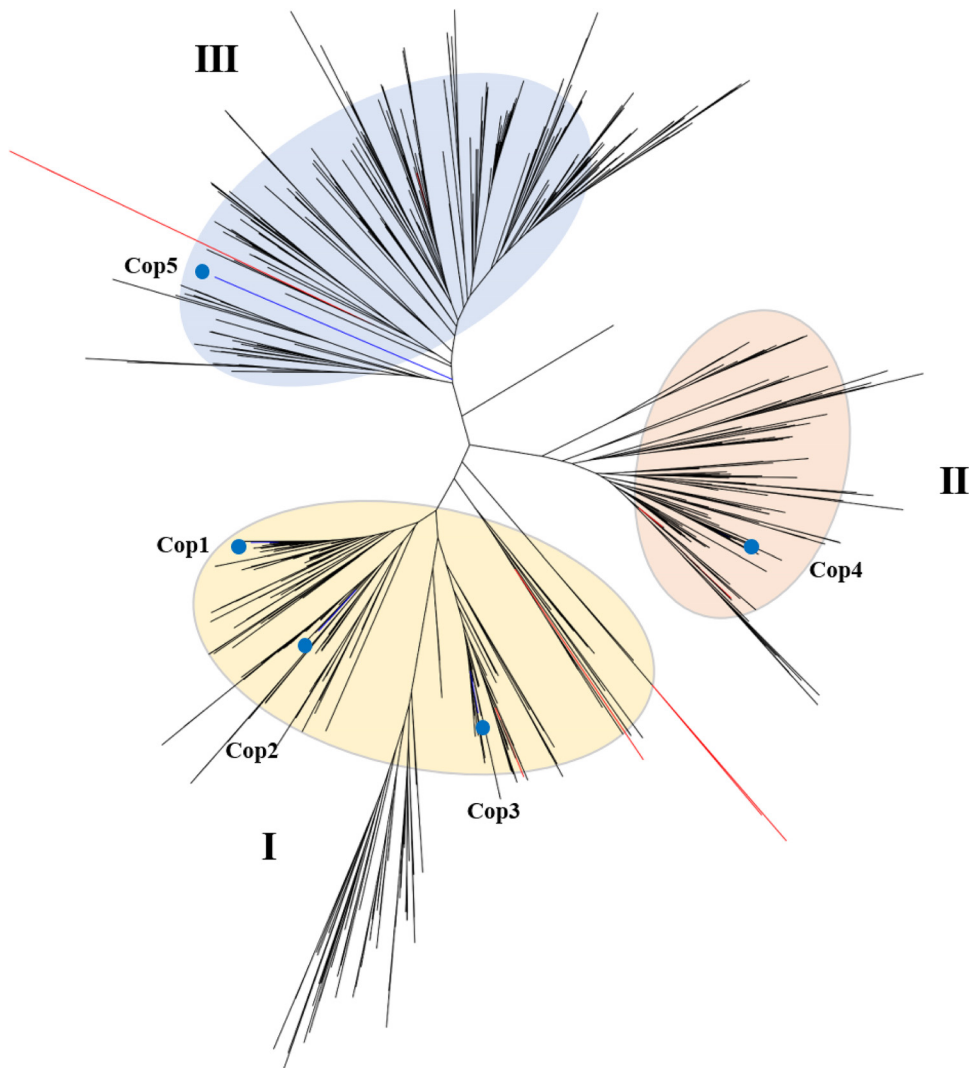
**FIG 3** Total CAZymes from *A. areolatum* and other fungal species. AA, auxiliary activities; GH, glycoside hydrolase; GT, glycosyltransferase; CBM, carbohydrate-binding module; PL, polysaccharide lyase. *L. echinatus*, *Lactarius echinatus*; *R. foetens*, *Russula foetens*; *L. deliciosus*, *Lactarius deliciosus*; *A. ellisii*, *Albatrellus ellisii*; *A. areolatum*, *Amylostereum areolatum*; *S. hirsutum*, *Stereum hirsutum*; *H. irregulare*, *Heterobasidion irregulare*; *Peniophora sp.*, *Peniophora sp.* strain CONT; *H. annosum*, *Heterobasidion annosum*; *H. parviporum*, *Heterobasidion parviporum*.

genes identified in *C. cinereus* from the 83 Agaricomycetes genomes. The *A. areolatum* genome contained 10 terpene synthase genes, consistent with the number of such enzymes in Agaricomycetes. A phylogenetic tree was constructed using the orthologs of three well-characterized *C. cinereus* terpene synthases, Cop3, Cop4, and Cop5, that were identified in the *A. areolatum* genome (Fig. 5). Terpene synthase genes of *A. areolatum* clustered with the Cop STS genes into three major groups, suggesting that the enzymes in each cluster might produce the related terpenoid compounds through conserved cyclization pathways (Fig. S3).

In fungi, secondary metabolite biosynthetic genes usually locate in contiguous clusters (24). Interestingly, terpene synthase gene clusters in clade II were conserved well across Russulales. Synteny among *A. areolatum*, *S. hirsutum*, and *H. irregulare* showed that genes involved in the synthesis of terpenoid compounds had good collinearity, indicating that Russulales might share conserved and core genes related to terpene metabolism (Fig. 6). The well-conserved gene cluster contained mevalonate kinase (MK), EGR12 (COG1557), nonplant terpene cyclases (cd00687), and enoyl-coenzyme A (enoyl-CoA) hydratase/isomerase and pkinase, in which some genes have been proven to be related to the mevalonate pathway (one of the pathways functioning to synthesize terpenoids). In addition to the core terpene cyclases involved in sesquiterpene metabolism, ABC transporter ATP-binding protein (CL0023) and P450 (PF00067) were also present in the sesquiterpene clusters. These enzymes might play



**FIG 4** Predicted phenotypic-category genes in *A. areolatum*. The distribution of phenotypic categories of *A. areolatum* gene orthologs was determined using the PHI database; percentages are based on 661 hits.



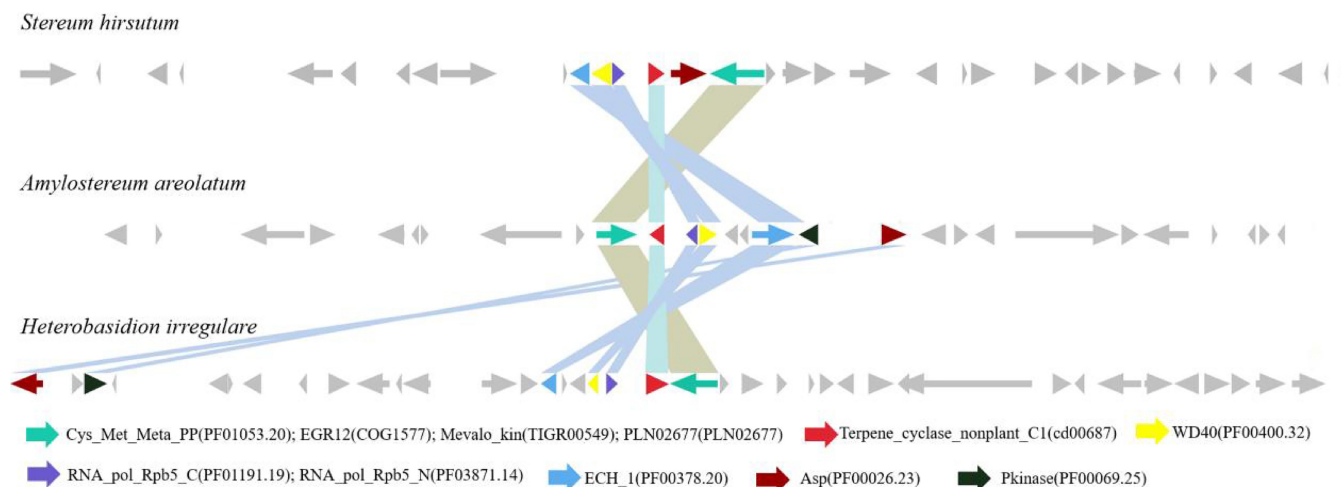
**FIG 5** Phylogram of terpene synthase homologs identified in 83 Agaricomycetes genomes. We used 902 putative terpene synthase genes from two orthologous groups, in which five sesquiterpene synthase homologs (named Cop1 to Cop5) identified in *C. cinereus* were included. The red and blue branches represent sesquiterpene synthase of *A. areolatum* and *C. cinereus*, respectively.

a role in modifications such as oxidation and hydroxylation, as well as in transportation of sesquiterpene.

## DISCUSSION

**General genomic features and phylogeny.** The genome assembly was about 53 Mb, comparable to the size of assemblies for the related members of Russulales, which range from 26 Mb to 122 Mb (26–30). The *A. areolatum* genome contained 15,611 coding genes, similar to the number of coding genes in Russulales fungus *S. hirsutum* (26) (14,066), *Peniophora* sp. strain CONT (30) (18,945), and *H. irregulare* (27) (13,275). This indicated that the genome assembly of *A. areolatum* was comparable in fragmentation and continuity to the other high-quality white rot fungal genomes. The phylogenetic tree showed that *A. areolatum* was evolutionally close to the plant-pathogenic fungus *Peniophora* sp. strain CONT rather than to *S. hirsutum* (fungi belonging to the same family as *A. areolatum*). Another study analyzing the phylogenetic relationship of Russulales species on the basis of sequences of the nuc ribosomal DNA (rDNA) ITS1-5.8S-ITS2 and D1-D2 domains of nuc 28S rDNA showed that *Amylostereum* belonged to the family Echinodontiaceae (31). This result may have been due to the





**FIG 6** Synteny of terpenoid gene cluster among *A. areolatum*, *S. hirsutum*, and *H. irregulare*.

limited availability of Russulales resources and the lack of genomes of species with relatively close evolutionary relationships. We would likely obtain more detailed results by performing studies with additional genomes of sibling species in the future.

**Protein families involved in degrading plant cell wall and cuticle.** *S. noctilio* females drill a series of test bores during oviposition to decide whether the trees are suitable for depositing eggs or not, basing on the perceived moisture content and resin pressure of the pine stems (2). All tunnels drilled by females are filled with venom mucus and fungus, but less than half of the tested-suitable tunnels contain eggs. The presence of venom and fungal symbiont in tunnels may cause weakening of hosts and dropping of resin pressure, conditions which provide a more suitable microenvironment for egg hatching and larval development of the woodwasp (2). Besides, the behavioral and morphological adaptations of the woodwasp indicate that *S. noctilio* larvae do not ingest xylem and that they use the fungi for external digestion of recalcitrant lignocellulosic compounds instead (14). *A. areolatum* can secrete a large amount of cellulase, hemicellulose, and ligninase, which can effectively degrade the plant cell wall components, providing nutrients (e.g., glucose, mannose, galactose, acetic acid, xylose, etc. [14, 32]) to the larvae of *S. noctilio*. Therefore, successful colonization and growth of *A. areolatum* play vital roles in the development of *S. noctilio*. Plant cell walls represent the main barrier against fungus invasions. CAZymes play a crucial role in the metabolism of glycoconjugates, oligosaccharides, and polysaccharides. They are responsible for the breakdown of cell wall components of hosts to establish a successful infection process. The depolymerizing abilities of the fungi are directly proportional to their ability to secrete a range of CAZymes (33). In our analysis, the number of CAZymes in *A. areolatum* was comparable to the number in forest pathogens and significantly higher than that in the traditional saprophytic mushrooms, indicating that *A. areolatum* can break through plant cell walls and successfully establish infection in a manner similar to that seen with other forest pathogens. It can weaken pine trees by destroying the structure of plant cells and provide nutrients for the growth of *S. noctilio* larvae.

Plant cell walls contain abundant cellulose. GHs are common enzymes that degrade cellulose, hemicellulose, and starch (34). They are involved in the hydrolysis of the glycosidic bond between or within carbohydrate molecules. A total of 219 GHs classified into 45 families were predicted in the *A. areolatum* genome. Genes encoding GH3 and GH5 class enzymes outnumbered those encoding other GH enzymes. Those were similar to the composition characteristics of GHs in wood-rotting fungi (white, brown, and soft rot fungi) (33). Many enzymes related to degradation of cellulose and xylan belong to the GH3 and GH5 families (35), indicating that this symbiotic fungus can

degrade these two substances. In addition, the fungal genome was found to be rich in genes encoding GH18 members, which catalyze the decomposition of chitin. Chitinase can degrade fungal cell walls and inhibit spore germination, mycelial growth, germ tube elongation, and other developmental activities (36, 37). Therefore, it is speculated that GH18 has played a role in the antagonism of other endophytic fungi such as *Ophiostoma minus* and *Chaetomium globosum* on Mongolian pine (15). Ninety percent of lignocellulose-degrading fungi contain genes encoding lytic polysaccharide mono-oxygenases (LPMOs) (38), which are classified into AA families in the CAZy database. These enzymes are mainly involved in the depolymerization of noncarbohydrate structural components (lignin) or are found as primary cell wall components of plants (34, 39). In this study, we detected 62 and 25 copies of AA3 and AA1 in *A. areolatum*, respectively. AA3 families (GMC oxidoreductases/dehydrogenases) are predominantly found in forest pathogens such as *S. hirsutum* (48 AA3 genes) and *Heterobasidion annosum* (32 AA3 genes) (33). The AA3 family contains enzymes from the glucose-methanol-choline (GMC) family of oxidoreductases, which assists in the activity of other AA family enzymes or supports the action of glycoside hydrolases in the degradation of lignocellulose via their reaction products (40). They promote the decomposition of lignocellulose by generating H<sub>2</sub>O<sub>2</sub> and recycling the electron donors and acceptors required for the oxidative attack of polymers (40).

Laccase is another important lignin-degrading enzyme belonging to the AA1 family. Catalytic oxidation of laccase is involved in the electron capturing of substrates, transfer of electrons between the copper ions in the active center, and, finally, transfer of electrons to O<sub>2</sub> and reduction of O<sub>2</sub> to water (41). In fungi, laccases have diverse roles, such as mediation of interactions between fungal pathogens and host plants, stress defense, morphogenesis, and lignin degradation (42, 43). In our study, the high numbers of AA3 and AA1 genes contributed to the oxidative degradation of *Pinus*, which gives *A. areolatum* a strong ability to penetrate plant cell walls and degrade lignin. This is not only good for fungus colonization and growth, it also provides nutrients to the wasp larvae.

**Virulence-associated genes.** The pathogen-host interactions database (PHI-base) catalogues more than 2,800 genes from fungi, bacteria, and protist pathogens, with experimentally verified pathogenicity, virulence, and effector genes (19, 44, 45). The inactivation or reduction of the expression of these genes can reduce or eliminate pathogenic ability (44). Among all the predicted genes obtained by searches against the PHI-base database, those associated with reduced virulence, unaffected pathogenicity, and loss of pathogenicity were the most frequent in *A. areolatum*. However, only a few virulence-associated genes (associated with increased virulence [hypervirulence] and effectors [plant avirulence determinants]) were identified. Physiological experiments have shown that inoculation with *A. areolatum* alone did not significantly reduce tree potential or affect tree growth (46). The characteristics of virulence-related genes of *A. areolatum* described in our analysis would help explain this phenomenon. Absence of expression or low expression of lethal genes can render the fungus unable to survive (44). There were only 9.23% lethal genes identified in the genome. This might have been because *A. areolatum* is a typical symbiotic fungus; the wasps would protect the fungus before introducing it into a suitable host substrate. Among the fungal transporters, ATP-binding cassette (ABC) transporters and the major facilitator super-family (MFS) are the two largest superfamilies. The ABC transporters are multicomponent, primarily active transporters, which transport both small molecules and macromolecules under conditions of ATP hydrolysis (47–49). They transport a broad range of compounds such as polysaccharides, drugs, sugars, heavy metals, oligopeptides, amino acids, and inorganic ions. Studies have shown that, in all sequenced fungi, *Schizosaccharomyces pombe* contained only 19 ABC proteins, which may be close to the minimal set for a free-living organism (50). *A. areolatum* has been found to have fewer ABC proteins than *S. pombe*, and, even more remarkably, it lacks the ABC-D transporters which have been found in every sequenced species except for *Encephalitozoon cuniculi*

and *S. pombe* (50). Only a few ABC proteins from fungi have been functionally characterized; the lack of ABC-D transporters in *A. areolatum* needs further study. We speculate that this may be related to the mutualisms of the wasp and fungus and the fact that *S. noctilio* is a secondary pest that usually colonizes relatively weaker pines (46). In comparison with free-living fungi, *A. areolatum* needs only to adapt to a relatively simple living environment; it grows inside the xylem of weaker pine stems and is transferred by its symbiotic insect hosts. In our results, the symbiotic fungus had a large number of carbohydrate enzyme genes but few virulence and transporter genes. This was consistent with their symbiotic relationship; *A. areolatum* provides nutrients for the growth of *S. noctilio* larvae by secreting extracellular enzymes, and *S. noctilio* protects and transfers the fungus before introducing into a new suitable host substrate by carrying the fungus inside a specialized organ, the mycangium.

**Sesquiterpene synthase clusters.** Female woodwasps are more likely to be attracted by trees that are weakened or that had been attacked previously (21). Volatile components are critical factors that are necessary for *S. noctilio* to locate suitable hosts and oviposition spots in different ranges (22). Studies have shown that mated female woodwasps were more attractive to the volatile components of *A. areolatum* than to those of the plant hosts; sesquiterpene compounds may play a role in this process (24). An analysis of sequence data from Russulales showed that sesquiterpene synthase homologs were widespread among these fungi. However, despite the preponderance of sesquiterpene synthase homologs in fungi, relatively little is known about their activities and biological functions. Cloning and characterization of sesquiterpene synthases of *C. cinereus* to the *Escherichia coli* expression vector revealed that  $\alpha$ -muurolene,  $\beta$ -elemene,  $\gamma$ -muurolene, germacrene D, and  $\delta$ -cadinene were produced by Cop3 cultures.  $\delta$ -Cadinene,  $\beta$ -cubebene, sativene,  $\beta$ -copaene, and cubebol were detected in the headspace of Cop4 cultures (24). We speculated that the same compounds would appear in the volatile profiles of *A. areolatum*. What sesquiterpene components are main factors that enable *S. noctilio* to locate the hosts have been attacked previously needs further study, and the results would be important in developing effective lures for this pest.

The *A. areolatum* genome is the first reported genome sequence of a woodwasp symbiotic fungus. The genomic resources presented here, including the genome sequences and annotations, as well as the detailed lists of cell wall-degrading enzymes, virulence-associated genes, and sesquiterpene synthase clusters show that the symbiotic fungus can release chemicals, attracting more female woodwasps to oviposit, and can degrade plant cell wall components by secreting cellulase, hemicellulose, and ligninase, which provide *S. noctilio* with a suitable environment and nutrients for the larval growth. These results lay a foundation for our understanding of the mechanism of this combined fungus-insect damaging system.

## MATERIALS AND METHODS

**Strains and culture conditions.** In 2017, *A. areolatum* was isolated from *S. noctilio* females collected from Mongolian pines (*P. sylvestris* var. *mongolica*) in Jun De Forest Farm (130°17'47"E, 47°12'11"N), Hei Longjiang Province, China. The fungus strain was confirmed as *A. areolatum* by morphological characteristics and molecular analysis in the internal transcribed spacer (ITS) and large subunit (LSU) regions as described previously by Wang et al. (15). The fungi were cultured on potato dextrose agar (PDA) and preserved at the Beijing Key Laboratory for the Control of Forest Pest, Beijing Forestry University, Beijing, China. For genomic DNA extraction, the fungi were cultured on PDA plates for 2 weeks at 25°C. Mycelia were collected in sterile tubes, washed with sterile water, immediately frozen with liquid nitrogen, and stored at -80°C until used.

**Genomic DNA extraction and sequencing.** The total genomic DNA of *A. areolatum* was extracted based on an improved cetyl trimethyl ammonium bromide (CTAB) procedure (51) and was sequenced using the Illumina HiSeq 4000 and PacBio Sequel platforms. For second-generation sequencing on the HiSeq 4000 platform, a large fragment library was prepared using a TruSeq DNA PCR-Free library prep kit (catalog no. FC-121-3001; Illumina, San Diego, CA, USA) with an average insertion size of 270 bp. The library was quantified using an Agilent 2100 bioanalyzer instrument (Agilent DNA 1000 reagents; Agilent, Santa Clara, CA, USA) and real-time quantitative PCR (RT-qPCR). The qualified libraries were amplified within the flow cell on the cBot instrument for cluster generation (HiSeq 4000 PE cluster kit; Illumina). The clustered flow cell was loaded onto a HiSeq 4000 sequencer for paired-end sequencing (HiSeq 4000 SBS kit; Illumina) with recommended read lengths of 150 bp. For third-generation sequencing on a PacBio

Sequel platform, genomic DNA samples were sheared to >10 kb using g-Tube (Covaris, Woburn, MA, USA). Libraries were prepared using SMRTbell template prep kit 1.0 (code 100-259-100; Pacific Biosciences, Menlo Park, CA, USA) with BluePippin size selection for fragments of >7 kb. The resulting libraries were sequenced on the PacBio Sequel platform at the Beijing Genomics Institute (Shenzhen, Guangdong, China).

**Genome assembly.** The Illumina sequencing produced 6,839 Mb of clean data from 7,784 Mb of raw data using SOAP *de novo* technology. The low-quality reads obtained by PacBio sequencing were filtered using SMRT Analysis v. 2.3.0 (52), and the subreads were subjected to two rounds of error correction using Falcon v. 0.3.0 and Proovread 2.12. The corrected PacBio reads were assembled with Celera Assembler 8.3 and Falcon v. 0.3.0. The Illumina sequences were used for k-mer analysis (Fig. S1) and correcting the PacBio assembly. To obtain a high-quality assembly, the sequence errors were corrected by GATK v1.6-13. Then, the contigs were combined into scaffolds using the SSPACE\_Basic v. 2.0 tool (53) and the gaps were closed using PBJELLY 15.8.24. Finally, we applied a redundant pipeline to identify and remove the heterozygous contigs and scaffolds, based on pairwise sequence similarity searches (54). The assembly candidate was evaluated by the use of BUSCO v 3.0.2 (55) with the data set from the fungus lineage.

**Gene prediction and annotation.** Protein-coding genes were predicted through a combination of *de novo* prediction and transcriptome-based prediction methods. For *ab initio* predictions, SNAP v. 2010-07-28 (56), Augustus v. 3.2.1 (57), and GeneMark ES v. 4.21 (58) were used to predict coding genes. Then, transcriptome sequencing (RNA-Seq) data (unpublished) were mapped to the assembly using Tophat v2.0.8 and the transcripts were assembled to gene models by the use of Cufflinks v2.1.1 (59). Finally, all gene models predicted by the methods named above were combined into a nonredundant set of gene structures by the use of EVIDENCEModeler (EVM) (60). The tRNA regions and secondary structures were detected using tRNAscan-SE v. 1.3.1 (61). The rRNAs were analyzed using RNAMmer software, and the small RNAs (sRNAs) were predicted using Infernal to search against the Rfam (v. 9.1) database (62, 63). The predicted gene models were functionally annotated by the use of BLASTp searches against the National Center for Biotechnology Information nonredundant database, Swiss-Prot (64), TrEMBL, Cluster of Orthologous Groups (COG) (65), Gene Ontology (GO) (66), and Kyoto Encyclopedia of Genes and Genomes (KEGG) (67, 68).

**Transposable elements and tandem repeat identification.** To evaluate the TEs within the *A. areolatum* genome, a sequence alignment prediction method and a *de novo* prediction method were used. For the sequence alignment prediction method, the TEs were searched with the supporting database Repbase using RepeatMasker v. 4-0-6 (<http://www.repeatmasker.org/>) (69). The optimized default parameters were utilized with the “-lib” option to find repeats, and the “-species Fungi” option was applied in a separate analysis to find fungal repeats. Repeat Protein Masker and the transposon protein library associated with RepeatMasker were also used to identify TEs. The *de novo* prediction method was used firstly to produce an eXtended Database Format (XDF) using buildXDFDatabase. Then, TE models were established using Repeat Modeler with the XDF database and these models were then used to predict TEs using Repeat Masker. Tandem repeats were evaluated using Tandem Repeat Finder v. 4.04 (<http://tandem.bu.edu/trf/trf.html>) (70).

**Phylogenetic analysis.** The genome sequences of 79 Agaricomycetes species (see Table S4 in the supplemental material) were downloaded from the NCBI database in FASTA format. A group of orthologous genes (phylogenetically conserved) in the fungal genome were obtained by the use of OrthoFinder 2.2.7 (71). We selected 41 single-copy orthologs to build the phylogenetic tree. Sequence alignment was done using MAFFT 7.409, and the conserved sites were extracted and concatenated by the use of Gblocks 0.91b. We ran ProtTest v2.0 to select the most appropriate model and used RAxML 8.2.12 (72) to build the ML gene tree with the “-f a -x 12345 -p 12345 -m PROTGAMMAILGF -N 1000” options (73). *Wallemia sebi* (GenBank accession number [GCF\\_000263375.1](https://www.ncbi.nlm.nih.gov/nuccore/GCF_000263375.1)) protein sequences were used as an outgroup (74, 75).

**CAZyme analysis.** CAZymes in *A. areolatum* and 10 other fungi were identified and annotated using BLAST (76), and the dbCAN annotation program HMMER 3 (<http://bcb.unl.edu/dbCAN2/index.php>) was used to search against the CAZy (carbohydrate-active enzyme) database (v. 2017-09; <http://www.cazy.org/>) (77, 78). The results were combined when the E value was less than 1E-05. The class II peroxidases and DyPs (dye-decolorizing peroxidases) were further confirmed by BLAST searches against the Peroxi-Base database (<http://peroxibase.toulouse.inra.fr/>) (79).

**Prediction of virulence-related factors.** Candidate virulence-associated genes were identified within the *A. areolatum* genome using BLASTp to search against PHI-base v. 4.3 (<http://www.phi-base.org/>) (20). Protein alignments were performed to identify putative virulence-associated genes in *A. areolatum* with more than 40% identity and 70% query coverage (80). The virulence factors (VFs) were searched using BLAST against the Virulence Factor Database (VFDB; <http://www.mgc.ac.cn/VFs/>) (81). The Transporter Classification Database (TCDB) contains sequences, classifications, and structural, functional, and evolutionary information about transport systems from a variety of taxa (82). Candidate transporters in *A. areolatum* were identified based on searches of the Transporter Classification Database (TCDB; <http://www.tcdb.org/>) with an E value threshold of 1E-05 and identity values of >40% (83). The sequence alignment of ABC transporters of fungi of Agaricomycotina was performed in MAFFT 7.409, and the phylogenetic tree was built using FastTree 2.1.7.

**Secondary metabolism gene prediction.** The secondary metabolism biosynthesis genes and gene clusters were predicted with AntiSMASH 5.0.0 (84) in the genomes of *A. areolatum*, *S. hirsutum*, *Peniophora* sp. strain CONT, and *H. irregulare*. Putative fungal terpene synthase sequences of 83 Agaricomycetes genomes were obtained in two ortholog groups. Five sesquiterpene synthase homo-

logues (named Cop1 to Cop5) identified in *C. cinereus* were also included. For phylogenetic tree construction, alignments were manually inspected to exclude sequences that seemed to be incorrectly annotated (e.g., sequences that appeared to be too short or too long). The sequence alignment of 902 terpene synthase genes was performed in MAFFT 7.409, and the phylogenetic tree was built using FastTree 2.1.7.

**Data availability.** The genome assembly has been deposited in the NCBI/DDJB/GenBank genome database under accession number [SAXG00000000](https://doi.org/10.1017/S0007485300009330). The *A. areolatum* raw sequence data have been deposited in the NCBI database under BioProject accession number [PRJNA513942](https://doi.org/10.1017/S0007485300009330) and BioSample accession number [SAMN10716400](https://doi.org/10.1017/S0007485300009330).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, TIF file, 0.1 MB.

**FIG S2**, TIF file, 1.5 MB.

**FIG S3**, TIF file, 0.1 MB.

**TABLE S1**, XLSX file, 0.01 MB.

**TABLE S2**, XLSX file, 0.01 MB.

**TABLE S3**, XLSX file, 0.02 MB.

**TABLE S4**, XLSX file, 0.03 MB.

**TABLE S5**, XLSX file, 0.01 MB.

**TABLE S6**, XLSX file, 0.01 MB.

**TABLE S7**, XLSX file, 0.01 MB.

## ACKNOWLEDGMENTS

This work was supported by the Beijing Science and Technology Planning Project (accession numbers Z191100008519004 and Z171100001417005) and Chinese National Natural Science Foundation (31870642) and National Key R & D Program of China (2018YFC1200400). These foundations provide assistance in DNA sequencing and computation infrastructure.

## REFERENCES

- Ryan K, Hurley BP. 2012. Life history and biology of *Sirex noctilio*. Springer Netherlands, Dordrecht, Netherlands.
- Slippers B, de Groot P, Wingfield MJ. 2012. The *Sirex* woodwasp and its fungal symbiont. Springer Netherlands, Dordrecht, Netherlands.
- Spradbery JP, Kirk AA. 1978. Aspects of the ecology of siricid woodwasps (Hymenoptera: Siricidae) in Europe, North Africa and Turkey with special reference to the biological control of *Sirex noctilio* F. in Australia. Bull Entomol Res 68:341–359. <https://doi.org/10.1017/S0007485300009330>.
- Wermelinger B, Thomsen IM. 2012. The woodwasp *Sirex noctilio* and its associated fungus *Amylostereum areolatum* in Europe. Springer Netherlands, Dordrecht, Netherlands.
- Hurley BP, Slippers B, Wingfield MJ. 2007. A comparison of control results for the alien invasive woodwasp, *Sirex noctilio*, in the southern hemisphere. Agric For Entomol 9:159–171. <https://doi.org/10.1111/j.1461-9563.2007.00340.x>.
- Groot PD, Nystrom K, Scarr T. 2006. Discovery of *Sirex noctilio* (Hymenoptera: Siricidae) in Ontario, Canada. Great Lakes Entomol 39: 49–53.
- Li D, Shi J, Lu M, Ren L, Zhen C, Luo Y. 2015. Detection and identification of the invasive *Sirex noctilio* (Hymenoptera: Siricidae) fungal symbiont, *Amylostereum areolatum* (Russulales: Amylostereaceae), in China and the stimulating effect of insect venom on laccase production by *A. areolatum* YQL03. J Econ Entomol 108:1136–1147. <https://doi.org/10.1093/jee/tov072>.
- Talbot P. 1977. The *Sirex*-*Amylostereum*-*Pinus* association. Annu Rev Phytopathol 15:41–54. <https://doi.org/10.1146/annurev.py.15.090177.000353>.
- Coutts MP. 1969. The mechanism of pathogenicity of *Sirex noctilio* on *Pinus radiata*. I. Effects of the symbiotic fungus *Amylostereum* sp. (Thelephoraceae). Aust J Bio Sci 22:915–924. <https://doi.org/10.1071/B19690915>.
- Hajek AE, Nielsen C, Kepler RM, Long SJ, Castrillo L. 2013. Fidelity among *Sirex* woodwasps and their fungal symbionts. Microb Ecol 65:753–762. <https://doi.org/10.1007/s00248-013-0218-z>.
- Kukor JJ, Martin MM. 1983. Acquisition of digestive enzymes by siricid woodwasps from their fungal symbiont. Science 220:1161–1163. <https://doi.org/10.1126/science.220.4602.1161>.
- Madden JL, Coutts MP. 1979. The role of fungi in the biology and ecology of woodwasps (Hymenoptera: Siricidae), p 165–174. In Batra LR (ed), Insect-fungus symbiosis: nutrition, mutualism and commensalism. Montclair, NJ, USA.
- Yousuf F, Carnegie AJ, Bashford R, Bedding RA, Nicol HI, Gurr GM. 2014. Bark beetle (*Ips grandicollis*) disruption of woodwasp (*Sirex noctilio*) biocontrol: direct and indirect mechanisms. For Ecol Manage 323: 98–104. <https://doi.org/10.1016/j.foreco.2014.03.009>.
- Thompson BM, Bodart J, McEwen C, Gruner DS. 2014. Adaptations for symbiont-mediated external digestion in *Sirex noctilio* (Hymenoptera: Siricidae). Ann Entom Soc Amer 107:453–460. <https://doi.org/10.1603/AN13128>.
- Wang L-X, Ren L-L, Liu X-B, Shi J, Wang J-Z, Luo Y-Q. 2019. Effects of endophytic fungi in Mongolian pine on the selection behavior of woodwasp (*Sirex noctilio*) and the growth of its fungal symbiont. Pest Manag Sci 75:492–505. <https://doi.org/10.1002/ps.5146>.
- Taylor KL. 1981. The *Sirex* woodwasp: ecology and control of an introduced forest insect, p 231–248. In Kitching RL, Jones RE (ed), The ecology of pests: some Australian case histories. CSIRO, Canberra, Australia.
- Gruben BS, Mäkelä MR, Kowalczyk JE, Zhou M, Benoit-Gelber I, Vries R. 2017. Expression-based clustering of CAZyme-encoding genes of *Aspergillus niger*. BMC Genomics 18:900. <https://doi.org/10.1186/s12864-017-4164-x>.
- van der Does HC, Rep M. 2007. Virulence genes and the evolution of host specificity in plant-pathogenic fungi. Mol Plant Microbe Interact 20: 1175–1182. <https://doi.org/10.1094/MPMI-20-10-1175>.
- Liu F, Ou XC, Zhan RL. 17 April 2018, posting date. Whole-genome sequence and genome annotation of *Xanthomonas citri* pv. *mangiferaeindicae*, causal agent of bacterial black spot on *Mangifera indica*. Arch Microbiol <https://doi.org/10.1007/s00203-018-1513-3>.
- Baldwin TK, Winnenburg R, Urban M, Rawlings C, Koehler J, Hammond-Kosack KE. 2006. The pathogen-host interactions database (PHI-base) provides insights into generic and novel themes of pathogenicity. Mol Plant Microbe Interact 19:1451–1462. <https://doi.org/10.1094/MPMI-19-1451>.

21. Fernández Ajó AA, Martínez AS, Villacide JM, Corley JC. 2015. Behavioural response of the woodwasp *Sirex noctilio* to volatile emissions of its fungal symbiont. *J Appl Entomol* 139:654–659. <https://doi.org/10.1111/jen.12211>.
22. Simpson R. 1976. Bioassay of pine oil components as attractants for *Sirex noctilio* (Hymenoptera: Siricidae) using electroantennogram techniques. *Entomol Exp Appl* 19:11–18. <https://doi.org/10.1111/j.1570-7458.1976.tb02576.x>.
23. Sarvary MA, Hajek AE, Böröczky K, Raguso RA, Cooperband MF. 2016. Investigating the effects of symbiotic fungi on the flight behaviour of *Sirex noctilio* (Hymenoptera: Siricidae). *Can Entomol* 148:543–551. <https://doi.org/10.4039/tce.2016.10>.
24. Agger S, Lopez-Gallego F, Schmidt-Dannert C. 2009. Diversity of sesquiterpene synthases in the basidiomycete *Coprinus cinereus*. *Mol Microbiol* 72:1181–1195. <https://doi.org/10.1111/j.1365-2958.2009.06717.x>.
25. Peiris D, Dunn WB, Brown M, Kell DB, Roy I, Hedger JN. 2008. Metabolite profiles of interacting mycelial fronts differ for pairings of the wood decay basidiomycete fungus *Stereum hirsutum* with its competitors *Coprinus micaceus* and *Coprinus disseminatus*. *Metabolomics* 4:52–62. <https://doi.org/10.1007/s11306-007-0100-4>.
26. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martínez AT, Otillar R, Spatafora JW, Yadav JS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de Vries RP, Ferreira P, Findley K, Foster B, Gaskell J, Glotzer D, Górecki P, Heitman J, Hesse C, Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Köhler A, Lumbs U, Kumar TKA, Kuo A, LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin R, McLaughlin DJ, Morgenstern I, Morin E, Murat C, Nagy LG, Nolan M, Ohm RA, Patyshakuliyeva A, et al. 2012. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336:1715–1719. <https://doi.org/10.1126/science.1221748>.
27. Sillo F, Garbelotto M, Friedman M, Gonther P. 2015. Comparative genomics of sibling fungal pathogenic taxa identifies adaptive evolution without divergence in pathogenicity genes or genomic structure. *Genome Biol Evol* 7:3190–3206. <https://doi.org/10.1093/gbe/evv209>.
28. Zeng Z, Sun H, Vainio EJ, Raffaello T, Kovalchuk A, Morin E, Duplessis S, Asiegbu FO. 2018. Intraspecific comparative genomics of isolates of the Norway spruce pathogen (*Heterobasidion parviporum*) and identification of its potential virulence factors. *BMC Genomics* 19:220. <https://doi.org/10.1186/s12864-018-4610-4>.
29. Li H, Wu S, Ma X, Chen W, Zhang J, Duan S, Gao Y, Kui L, Huang W, Wu P, Shi R, Li Y, Wang Y, Li J, Guo X, Luo X, Li Q, Xiong C, Liu H, Gui M, Sheng J, Dong Y. 2018. The genome sequences of 90 mushrooms. *Sci Rep* 8:9982. <https://doi.org/10.1038/s41598-018-28303-2>.
30. Nagy LG, Riley R, Tritt A, Adam C, Daum C, Floudas D, Sun H, Yadav JS, Pangilinan J, Larsson K-H, Matsuura K, Barry K, Labutti K, Kuo R, Ohm RA, Bhattacharya SS, Shirouzu T, Yoshinaga Y, Martin FM, Grigoriev IV, Hibbett DS. 2016. Comparative genomics of early-diverging mushroom-forming fungi provides insights into the origins of lignocellulose decay capabilities. *Mol Biol Evol* 33:959–970. <https://doi.org/10.1093/molbev/msv337>.
31. Liu SL, Zhao Y, Dai YC, Nakasone KK, He SH. 2017. Phylogeny and taxonomy of Echinodontium and related genera. *Mycologia* 109:1–10. <https://doi.org/10.1080/00275514.2017.1369830>.
32. Kirk TK, Cullen D. 1998. Enzymology and molecular genetics of wood degradation by white-rot fungi, p 273–308. *In* Young RA, Akhtar M (ed), *Environmentally friendly technologies for the pulp and paper industry*. John Wiley & Sons, Inc, New York, NY.
33. Kameshwar AKS, Qin W. 2018. Comparative study of genome-wide plant biomass-degrading CAZymes in white rot, brown rot and soft rot fungi. *Mycology* 9:93–13. <https://doi.org/10.1080/21501203.2017.1419296>.
34. Park Y-J, Jeong Y-U, Kong W-S. 2018. Genome sequencing and carbohydrate-active enzyme (CAZyme) repertoire of the white rot fungus *Flammulina elastica*. *Int J Mol Sci* 19:2379. <https://doi.org/10.3390/ijms19082379>.
35. Nguyen STC, Freund HL, Kasanjian J, Berlemont R. 2018. Function, distribution, and annotation of characterized cellulases, xylanases, and chitinases from CAZy. *Appl Microbiol Biotechnol* 102:1629–1637. <https://doi.org/10.1007/s00253-018-8778-y>.
36. Mauch F, Mauch-Mani B, Boller T. 1988. Antifungal hydrolases in pea tissue: II. Inhibition of fungal growth by combinations of chitinase and beta-1,3-glucanase. *Plant Physiol* 88:936–942. <https://doi.org/10.1104/pp.88.3.936>.
37. Arlorio M, Ludwig A, Boller T, Bonfante P. 1992. Inhibition of fungal growth by plant chitinases and  $\beta$ -1,3-glucanases. *Protoplasma* 171:34–43. <https://doi.org/10.1007/BF01379278>.
38. Kracher D, Scheiblbrandner S, Felice AK, Breslmayr E, Preims M, Ludwicka K, Haltrich D, Eijsink VG, Ludwig R. 2016. Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science* 352:1098–1101. <https://doi.org/10.1126/science.aaf3165>.
39. Lévassieur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* 6:41. <https://doi.org/10.1186/1754-6834-6-41>.
40. Qin X, Su X, Luo H, Ma R, Yao B, Ma F. 2018. Deciphering lignocellulose deconstruction by the white rot fungus *Irpex lacteus* based on genomic and transcriptomic analyses. *Biotechnol Biofuels* 11:58. <https://doi.org/10.1186/s13068-018-1060-9>.
41. Ma Y, Li L, Tian H, Lu M, Megharaj M, He W. 2019. Transcriptional analysis of the laccase-like gene from *Burkholderia cepacia* BNS and expression in *Escherichia coli*. *Appl Microbiol Biotechnol* 103:747–760. <https://doi.org/10.1007/s00253-018-9468-5>.
42. Wei W, Fang L, Yuji J, Guangmei W, Lixian G, Renliang C, Bingzhi C, Yuanping L, Yucheng D, Baogui X. 2015. The multigene family of fungal laccases and their expression in the white rot basidiomycete *Flammulina velutipes*. *Gene* 563:142–149. <https://doi.org/10.1016/j.gene.2015.03.020>.
43. Kameshwar AKS, Barber R, Qin W. 2018. Comparative modeling and molecular docking analysis of white, brown and soft rot fungal laccases using lignin model compounds for understanding the structural and functional properties of laccases. *J Mol Graph Model* 79:15–26. <https://doi.org/10.1016/j.jmgm.2017.10.019>.
44. Winnenburg R, Baldwin TK, Urban M, Rawlings C, Köhler J, Hammond-Kosack KE. 2006. PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res* 34(Database issue):D459–D464. <https://doi.org/10.1093/nar/gkj047>.
45. Triplett LR, Hamilton JP, Buell CR, Tisserat NA, Verdier V, Zink F, Leach JE. 2011. Genomic analysis of *Xanthomonas oryzae* isolates from rice grown in the United States reveals substantial divergence from known *X. oryzae* pathovars. *Appl Environ Microbiol* 77:3930–3937. <https://doi.org/10.1128/AEM.00028-11>.
46. Bashford R, Madden JL. 2012. The use of kairomone lures for the detection of *Sirex noctilio* in susceptible *Pinus radiata* plantations in Australia, p 159–166. *In* Slippers B, de Groot P, Wingfield MJ (ed), *The sirex woodwasp and its fungal symbiont*. Springer, Dordrecht, Netherlands.
47. Stergiopoulos I, Zwiers L-H, De Waard MA. 2002. Secretion of natural and synthetic toxic compounds from filamentous fungi by membrane transporters of the ATP-binding cassette and major facilitator superfamily. *Eur J Plant Pathol* 108:719–734. <https://doi.org/10.1023/A:1020604716500>.
48. Dimroth P. 1997. Primary sodium ion translocating enzymes. *Biochim Biophys Acta* 1318:11–51. [https://doi.org/10.1016/s0005-2728\(96\)00127-2](https://doi.org/10.1016/s0005-2728(96)00127-2).
49. Paulsen IT, Brown MH, Skurray RA. 1996. Proton-dependent multidrug efflux systems. *Microbiol Rev* 60:575–608. <https://doi.org/10.1128/MMBR.60.4.575-608.1996>.
50. Kovalchuk A, Driessen AJ. 2010. Phylogenetic analysis of fungal ABC transporters. *BMC Genomics* 11:177. <https://doi.org/10.1186/1471-2164-11-177>.
51. Watanabe M, Lee K, Goto K, Kumagai S, Sugita-Konishi Y, Hara-Kudo Y. 2010. Rapid and effective DNA extraction method with bead grinding for a large amount of fungal DNA. *J Food Prot* 73:1077–1084. <https://doi.org/10.4315/0362-028x-73.6.1077>.
52. Cloney R. 2016. DNA methylation: a SMRT analysis of prokaryotic epigenomes. *Nat Rev Genet* 17:195–195. <https://doi.org/10.1038/nrg.2016.22>.
53. Boetzer M, Henkel CV, Janzen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
54. Pruszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res* 44:e113. <https://doi.org/10.1093/nar/gkw294>.
55. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
56. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PIW. 2008. SNAP: a Web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24:2938–2939. <https://doi.org/10.1093/bioinformatics/btn564>.
57. Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene predic-

- tion method employing protein multiple sequence alignments. *Bioinformatics* 27:757–763. <https://doi.org/10.1093/bioinformatics/btr010>.
58. Borodovsky M, Lomsadze A. 2011. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatics* 35: 4.6.1–4.6.10. <https://doi.org/10.1002/0471250953.bi0406s35>.
  59. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7:562–578. <https://doi.org/10.1038/nprot.2012.016>.
  60. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9:R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
  61. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.955>.
  62. Lagesen K, Hallin P, Rodland E, Staerfeldt H, Rognes T, Ussery D. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>.
  63. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. 2009. Rfam: updates to the RNA families database. *Nucleic Acids Res* 37: D136–D140. <https://doi.org/10.1093/nar/gkn766>.
  64. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31:365–370. <https://doi.org/10.1093/nar/gkg095>.
  65. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36. <https://doi.org/10.1093/nar/28.1.33>.
  66. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29. <https://doi.org/10.1038/75556>.
  67. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 34(Database issue):D354–D357. <https://doi.org/10.1093/nar/gkj102>.
  68. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32(Database issue):D277–D280. <https://doi.org/10.1093/nar/gkh063>.
  69. Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6:11. <https://doi.org/10.1186/s13100-015-0041-9>.
  70. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580. <https://doi.org/10.1093/nar/27.2.573>.
  71. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16:157. <https://doi.org/10.1186/s13059-015-0721-2>.
  72. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
  73. Min B, Kim S, Oh Y-L, Kong W-S, Park H, Cho H, Jang K-Y, Kim J-G, Choi I-G. 2018. Genomic discovery of the hypsin gene and biosynthetic pathways for terpenoids in *Hypsizygus marmoreus*. *BMC Genomics* 19: 789. <https://doi.org/10.1186/s12864-018-5159-y>.
  74. Hibbett DS, Bauer R, Binder M, Giachini A, Hosaka K, Justo A, Larsson E, Larsson K-H, Lawrey JD, Miettinen O. 2014. 14 Agaricomycetes, p 373–429. *In* Systematics and evolution. Springer Netherlands, Dordrecht, Netherlands.
  75. Padamsee M, Kumar TKA, Riley R, Binder M, Boyd A, Calvo AM, Furukawa K, Hesse C, Hohmann S, James TY, LaButti K, Lapidus A, Lindquist E, Lucas S, Miller K, Shantappa S, Grigoriev IV, Hibbett DS, McLaughlin DJ, Spatafora JW, Aime MC. 2012. The genome of the xerotolerant mold *Wallenia sebi* reveals adaptations to osmotic stress and suggests cryptic sexual reproduction. *Fungal Genet Biol* 49:217–226. <https://doi.org/10.1016/j.fgb.2012.01.007>.
  76. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
  77. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a Web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40:W445–W451. <https://doi.org/10.1093/nar/gks479>.
  78. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 37(Database issue): D233–D238. <https://doi.org/10.1093/nar/gkn663>.
  79. Passardi F, Theiler G, Zamocky M, Cosio C, Rouhier N, Teixeira F, Margis-Pinheiro M, Ioannidis V, Penel C, Falquet L, Dunand C. 2007. PeroxiBase: the peroxidase database. *Phytochemistry* 68:1605–1611. <https://doi.org/10.1016/j.phytochem.2007.04.005>.
  80. Hittalmani S, Mahesh HB, Mahadevaiah C, Prasannakumar MK. 2016. De novo genome assembly and annotation of rice sheath rot fungus *Sarocladium oryzae* reveals genes involved in helvolic acid and cerulenin biosynthesis pathways. *BMC Genomics* 17:271. <https://doi.org/10.1186/s12864-016-2599-0>.
  81. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 33(Database issue):D325–D328. <https://doi.org/10.1093/nar/gki008>.
  82. Saier MH, Jr, Tran CV, Barabote RD. 2006. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* 34(Database issue):D181–D186. <https://doi.org/10.1093/nar/gkj001>.
  83. Liang X, Wang B, Dong Q, Li L, Rollins JA, Zhang R, Sun G. 2018. Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. *PLoS One* 13:e0196303. <https://doi.org/10.1371/journal.pone.0196303>.
  84. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.