

RESEARCH

Open Access



# Genome sequencing of *Elaeocarpus* spp. stem blight pathogen *Pseudocryphonectria elaeocarpicola* reveals potential adaptations to colonize woody bark

Yuchen Yang<sup>1</sup>, Dianguang Xiong<sup>1\*</sup>, Danyang Zhao<sup>2</sup>, Huayi Huang<sup>2\*</sup> and Chengming Tian<sup>1</sup>

## Abstract

**Background** *Elaeocarpus* spp. stem blight, caused by *Pseudocryphonectria elaeocarpicola*, is a destructive disease, which will significantly reduce the productivity and longevity of *Elaeocarpus* spp. plants, especially in the Guangdong Province of China. However, few information is available for *P. elaeocarpicola*. To unravel the potential adaptation mechanism of stem adaptation, the whole genome of *P. elaeocarpicola* was sequenced by using the DNBSAQ and PacBio platforms.

**Results** *P. elaeocarpicola* harbors 44.49 Mb genome with 10,894 predicted coding genes. Genome analysis revealed that the *P. elaeocarpicola* genome encodes a plethora of pathogenicity-related genes. Analysis of carbohydrate-active enzymes (CAZymes) revealed a rich variety of enzymes participated in plant cell wall degradation, which could effectively degrade cellulose, hemicellulose and xyloglucans in the plant cell wall and promote the invasion of the host plant. There are 213 CAZyme families found in *P. elaeocarpicola*, among which glycoside hydrolase (GH) family has the largest number, far exceeding other tested fungi by 53%. Besides, *P. elaeocarpicola* has twice as many genes encoding chitin and cellulose degradation as *Cryphonectria parasitica*, which belong to the same family. The predicted typical secreted proteins of *P. elaeocarpicola* are numerous and functional, including many known virulence effector factors, indicating that *P. elaeocarpicola* has great potential to secrete virulence effectors to promote pathogenicity on host plants. AntiSMASH revealed that the genome encoded 61 secondary metabolic gene clusters including 86 secondary metabolic core genes which was much higher than *C. parasitica* (49). Among them, two gene cluster of *P. elaeocarpicola*, cluster12 and cluster52 showed 100% similarity with the mycotoxins synthesis clusters from *Aspergillus steynii* and *Alternaria alternata*, respectively. In addition, we annotated cytochrome P450 related enzymes, transporters, and transcription factors in *P. elaeocarpicola*, which are important virulence determinants of pathogenic fungi.

**Conclusions** Taken together, our study represents the first genome assembly for *P. elaeocarpicola* and reveals the key virulence factors in the pathogenic process of *P. elaeocarpicola*, which will promote our understanding of its pathogenic mechanism. The acquired knowledge lays a foundation for further exploration of molecular interactions with the host and provide target for management strategies in future research.

\*Correspondence:

Dianguang Xiong  
xiongdianguang@126.com  
Huayi Huang  
huanghy@sinogaf.cn

Full list of author information is available at the end of the article



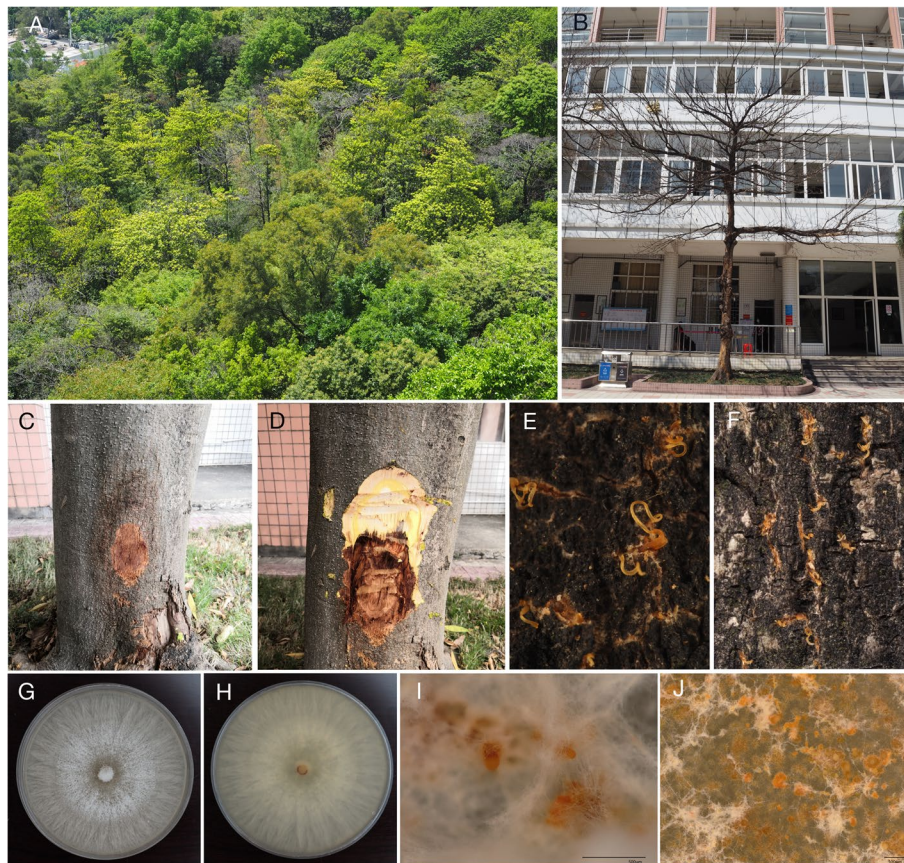
© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Keywords** Fungal genomic, *Pseudocryphonectria elaeocarpicola*, Pathogenic factors, CAZymes, Secondary metabolism, Stem blight

## Background

Blight and canker diseases have caused significant adverse effects on the production and management of woody plants worldwide. Among numerous pathogenic fungi, Cryptomeriaceae, a pathogenic family within Diaporthales, contains a variety of pathogenic fungi that cause blight and canker diseases [1, 2]. The most notorious species is *Cryphonectria parasitica* which causes chestnut blight throughout the world, reducing the longevity and productivity of infected trees [3]. *Chrysosporthe cubensis*, *Chrysosporthe deuterocubensis*, and *Chrysosporthe austroafricana* are the causal agent of eucalypt (*Eucalyptus* spp.) canker diseases in South America, Asia and Africa, respectively [4]. Recently, *Pseudocryphonectria elaeocarpicola* is found to cause severe stem blight on *Elaeocarpus* spp. in China [5].

*Elaeocarpus* spp. is an important garden tree species with significant economic and ecological value in Guangdong Province, China. However, *Elaeocarpus* spp. stem blight disease caused by *P. elaeocarpicola* is a restraint to the growth of *Elaeocarpus* spp., which threaten the landscape greening in Guangdong Province seriously (Fig. 1A, B). The disease mainly develops on areas of the trunk or branches of the host (Fig. 1C). The bark of infected branches appears scorched and withered (Fig. 1D). Besides, *P. elaeocarpicola* forms orange conidial tendrils on the host bark at the late stages of the disease cycle (Fig. 1E-J). Moreover, *P. elaeocarpicola* significantly shorten the lifespan and productivity of the infected trees, and most of affected trees died within five months [5]. However, the disease cycle, epidemiology, histocytology and molecular mechanism underlying the pathogenicity of *Elaeocarpus* spp. stem blight disease



**Fig. 1** Symptoms caused by *Pseudocryphonectria elaeocarpicola* on *Elaeocarpus* trees. **A, B** Dead trees; **C, D** Symptom of *P. elaeocarpicola* on *Elaeocarpus* bark; **E, F** Orange conidial tendrils formed on the cankered barks; **G-J** Morphology of *P. elaeocarpicola* from PDA

are scarcely reported. Therefore, predicting the virulence factors of *P. elaeocarpicola* through whole genome sequencing is necessary to study the pathogenic mechanism and develop the more effective disease management strategies.

The rapidly developing new generation sequencing technology provides high-quality, low-cost genomic data that can be used for diverse analyses. To date, genomic research has been performed on hundreds of multiple plant pathogenic fungi [6–8]. Genomic research has greatly contributed to the prediction of pathogenic fungal virulence factors and the analysis of invasion mechanism [9]. In addition, many bioinformatics computational tools and well-established fungal databases provide greater assistance in mining virulence factors in pathogenic fungal genome sequences. The commonly used analysis tools and databases are as follows: CAZydatabase, antiSMASH, Cytochrome P450 Engineering Database, SignalP, EffectorP and Transporter Classification Database [10–14]. With the help of these tools and databases, researchers are able to study the interactions between pathogenic fungi and hosts, thus providing more precise research objectives for the study of pathogenic fungal pathogenesis and control management.

The disease establishment of pathogenic fungi on host plants depends on diverse virulence factors, including cell wall degrading enzymes, effector proteins, secondary metabolites, transporter proteins and so on [15]. Pathogenic fungi require complete and complex gene regulatory networks to regulate these virulence factors to accomplish host penetration, nutrient acquisition, prevention or inhibition of host defense, and even kill plant tissue. Therefore, virulence factors are essential for the interaction between fungal pathogens and host. The carbohydrate-active enzymes (CAZymes) are crucial factors, which secreted by plant pathogenic fungi in the initial infection process [16–18]. CAZymes can break down the components of plant cell wall (cellulose, hemicellulose, xylan, xyloglucan, mannan and pectin), which enable pathogens to colonize host plants [19]. Therefore, these enzymes are particularly important during the early stages of infection for plant pathogenic fungi as well as the late infection stages [20]. Although the enzymatic activity is critical for the role of CAZymes, studies of *Botrytis cinerea*, *Sclerotinia sclerotiorum*, and *Mycosphaerella graminicola* have shown that CAZymes have a positive relationship with virulence and act as virulence factors [21–24]. After colonization on host plants, pathogens attack plant cells through a variety of strategies including secondary metabolites (SMs), cytochrome P450 monooxygenases and transporters [25–28].

Fungal pathogens produce a range of SMs with strong biological activities during the secondary metabolism,

which play important roles in the development and pathogenesis. The genes encoding SMs are usually clustered together [29], and this chromosomal architecture has facilitated the prediction and mining of secondary metabolism gene clusters in fungal genomes. Secondary metabolism gene clusters are minimally composed of backbone genes and tailoring enzymes genes. Backbone genes are the most central and critical genes in gene clusters, mainly including non-ribosomal peptide synthetase (NRPS) genes, terpene synthase genes and polyketide synthase (PKS) genes. Tailoring enzymes mainly include methyltransferases, cytochrome P450 monooxygenases, hydroxylases and epimerases [29]. Mycotoxins synthesized under the encode of secondary metabolism gene clusters and core genes have been proved to play key roles in pathogenic processes and can directly damage or kill host plant tissues. The well-known mycotoxin include T-toxin, deoxynivalenol (DON), ochratoxin A (OTA), aflatoxin and so on [30–33]. In addition to attacking host plants directly, secondary metabolites can also inhibit host immune defense mechanisms. Cytochrome P450 related enzyme (CYPs or P450) are also involved in the synthesis of secondary metabolites [34, 35]. CYPs are proteins containing heme and one of the largest protein families, present in all biological kingdoms [36]. Moreover, fungal CYPs also participate in multiple biological processes, including primary metabolism, promoting fungal adaptations to various ecological niches and denitrification. Notably, CYPs play key catalytic abilities in the biosynthesis of secondary metabolites. The reactions catalyzed by CYPs greatly enriches the chemical structure and biological activity of secondary metabolites [36].

Besides, transport proteins are a large group of membrane proteins that mediate the exchange of chemicals as well as signals inside and outside biological membranes. Meanwhile, active efflux by major facilitator superfamily (MFS) transporters or ATP-binding cassette (ABC) can develop resistance to various toxic compounds such as SMs, antibiotics and fungicides [37–39]. Furthermore, in order to survive in a wide variety of environments, fungi often use extracellular proteins to decompose nutrients, obtain metal ions, and mediate interactions with the host [40]. Notably, pathogens have the ability to secrete enzymes and other proteins to overcome the host defense responses or directly kill the host [41, 42]. Further correspondence studies based on the predicted results of these virulence factors could study the pathogenic mechanism of novel plant pathogens more effectively and provide ideas and directions for further development of prevention and control strategies.

In this study, we sequenced the genome of the *P. elaeocarpicola*, which caused *Elaeocarpus* spp. stem blight. Comparative analyses showed that the *P. elaeocarpicola*

genome encodes a variety of functional genes, including a lot of potential virulence factors. These virulence factors most likely play important functions in the invasion and colonization of host plants by *P. elaeocarpicola*. Notably, gene families related to cellulase, hemicellulase, and secondary metabolic are expanded in *P. elaeocarpicola*. The analysis of CAZymes shows that *P. elaeocarpicola* has a strong potential to secrete cell wall degrading enzymes. Meanwhile, the secondary metabolic genes in *P. elaeocarpicola* have great potential to play a dominant role in the biological processes of development and virulence. In addition, a lot of toxic effector proteins, CYPs and transcription factors with known homologous pathogenic effects are annotated in *P. elaeocarpicola* genome. In summary, the analysis results of *P. elaeocarpicola* genome reported in this article provide valuable information for studying the pathogenesis of this novel blight pathogen.

## Results

### Genome characteristics of *P. elaeocarpicola*

Genome of *P. elaeocarpicola*, a blight pathogen isolated from *Elaeocarpus* spp., was sequenced de novo and assembled to a genome size of 44.49 Mb. The genome sequence contained 23 scaffolds and coverage of 98.2% (Table 1; Fig. 2). The length of N50 was 3.42 Mb, and the largest scaffold was 8.5 Mb, the smallest scaffold was 4291 bp. The GC content was 53.18%, and length of the repetitive sequences in *P. elaeocarpicola* genome was 3.66 Mb, representing 8.72% of the genome (Table 1; Fig. 2). In addition, a total of 10,894 protein-coding genes were predicted in the *P. elaeocarpicola* genome. The average gene length of the *P. elaeocarpicola* genome is 1775 bp, with an intron length of 145 bp, which is similar to

other Ascomycota species [28, 43]. The *P. elaeocarpicola* genome size is comparable to other filamentous fungi such as *Valsa mali* (44.7 Mb), *C. parasitica* (43.9 Mb) and *Magnaporthe oryzae* (42.7 Mb).

Furthermore, to explore the evolutionary relationship and gene family of *P. elaeocarpicola*, we clustered the proteomes of *P. elaeocarpicola* with those of 12 typical stem and foliar fungi with different lifestyles using Orthofinder, and constructed a phylogenetic tree based on the results of gene family analysis (Fig. 3). The results of the ortholog family analysis showed that a total of 2925 single-copy orthologs were conserved in all analyzed fungi. There are 4882 genes that can be clustered into gene families in *P. elaeocarpicola*, and the number of gene families is 51. In addition, a total of 25 species-specific proteins were identified in *P. elaeocarpicola*, distributed among 10 unique gene families. Besides, the resulting phylogenomic tree shows that *P. elaeocarpicola* is closely related to *C. parasitica*, which matches the previous reports [5]. In addition, we performed synteny analysis of *P. elaeocarpicola* with *C. parasitica*. Synteny analysis revealed higher sequence similarity between *P. elaeocarpicola* and *C. parasitica* (Fig. 2).

### Genome functional annotation of *P. elaeocarpicola*

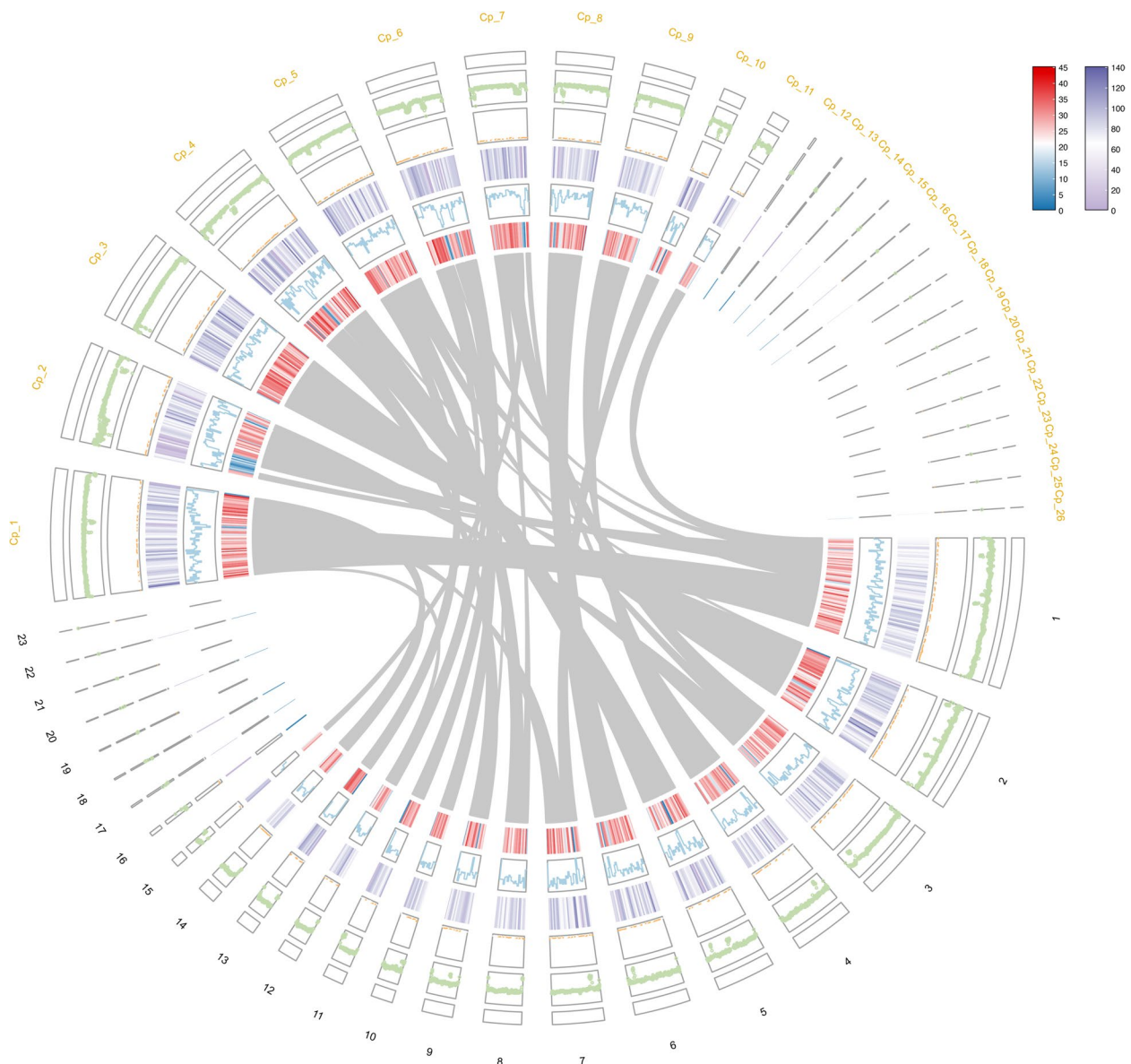
To further explore the function of the *P. elaeocarpicola*, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) database were used to predict the annotations of genes in *P. elaeocarpicola* (Fig. 4).

A total of 6011 genes were annotated in the GO database, which can be categorized into three subcategories based on the function of each gene: biological processes (19 branches), cellular components (3 branches), and molecular functions (13 branches; Fig. 4A). In the biological process subcategories, the three largest branches are metabolic, metabolic process (2882), cellular process (2873), and localization (820) are the three largest branches. Among the cellular component subcategories, the proportion of cellular anatomical entities (1846) is the highest. In the molecular function subcategories, catalytic activity (2975), binding (2779), and transporter activity (363) are ranked in the top three, respectively.

The results of KEGG pathway mapping shows that a total of 4463 genes were correspond to 46 metabolic pathways (Fig. 4B). The main pathways, in descending order of the number of genes involved, are as follows: MAPK signaling pathway – yeast (ko04011; 65 genes; Environmental Information Processing; Signal transduction), mTOR signaling pathway (ko04150; 44 genes; Environmental Information Processing; Signal transduction); ribosome (ko03010; 111 genes; Genetic Information Processing; Translation), spliceosome (ko03040; 96 genes; Genetic Information Processing; Translation);

**Table 1** Genome features of the *Pseudocryphonectria elaeocarpicola*

Feature	<i>P. elaeocarpicola</i>
Number of scaffolds	23
Coverage (%)	98.2
Length of genome assembly (Mbp)	44.49
Genome depth	28.78
N50(Mbp)	3.42
GC content (%)	53.18
Repeat size (Mbp)	3.66
Repeat rate (%)	8.7155
Gene number	10894
Average gene length (bp)	1775.16
Average exon size (bp)	516.4
Average coding sequence size (bp)	1499.61
Average intron size (bp)	144.72
ncRNA	1242

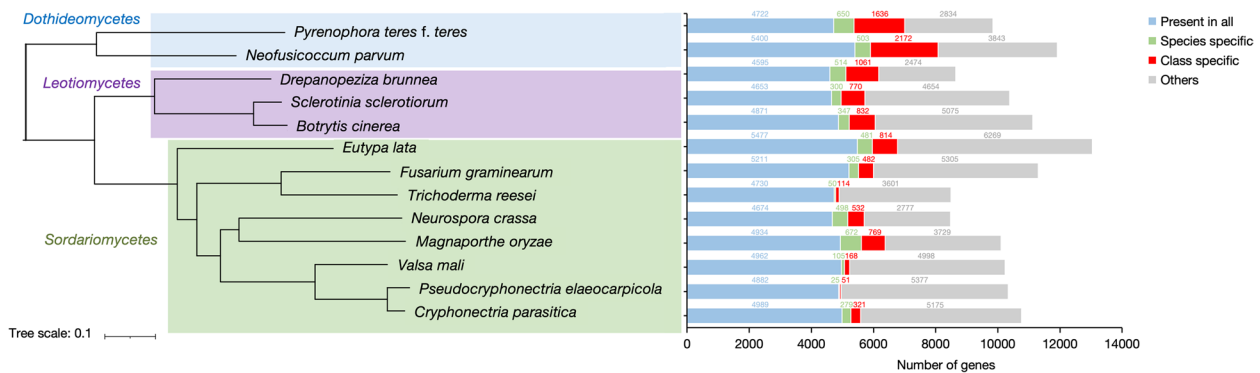


**Fig. 2** Circos-plot of genome features of *P. elaeocarpicola*. The circles from outside going in represent genomic synteny and comparison between *P. elaeocarpicola* and *Cryphonectria parasitica*, gene density, CDS density, GC skew, GC ratio, 1–23 represent *P. elaeocarpicola* chromosome skeleton, Cp\_1-Cp\_26 represent *C. parasitica* chromosome skeleton

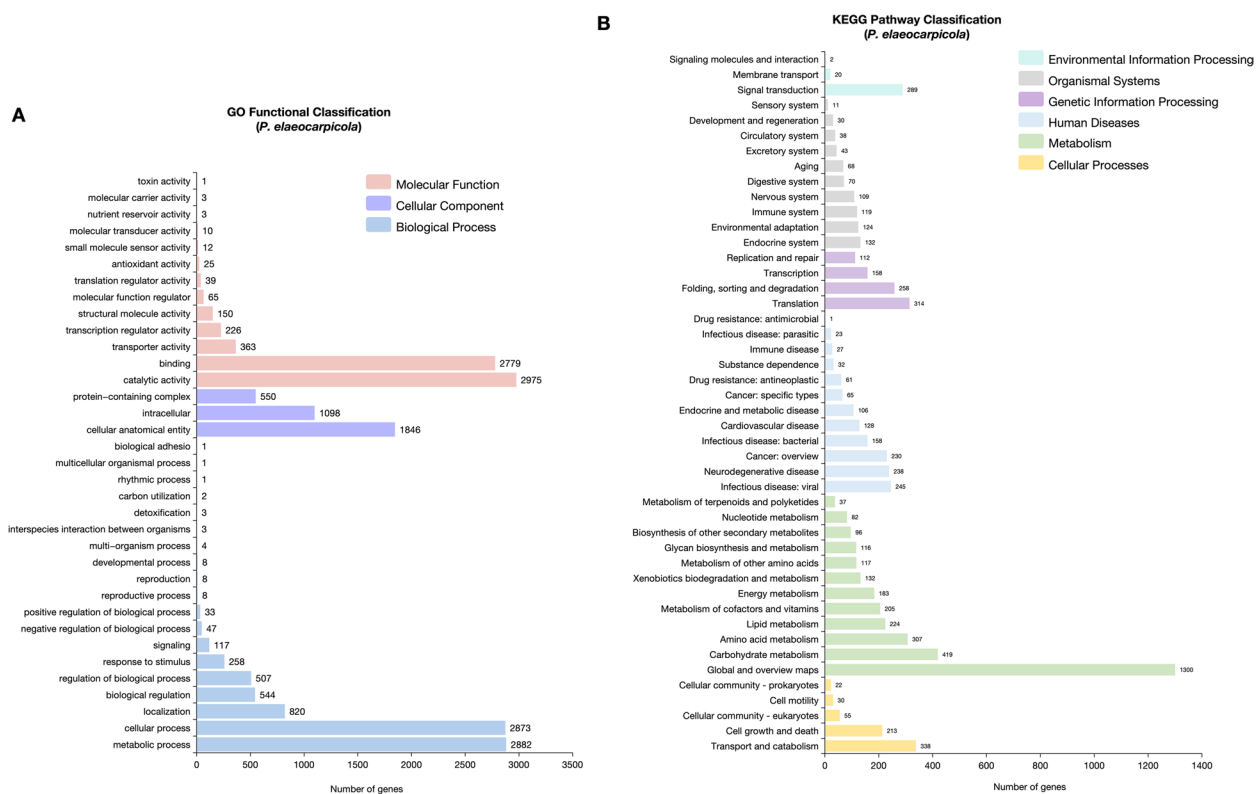
amyotrophic lateral sclerosis (ko05014; 158 genes; Human Diseases; Neurodegenerative disease), Pathways of neurodegeneration - multiple diseases (ko05022; 155 genes; Human Diseases; Neurodegenerative disease); metabolic pathways (ko01100; 1260 genes; Metabolism; Global and overview maps ), biosynthesis of secondary metabolites (ko01100; 465 genes; Metabolism; Global and overview maps ); cell cycle - yeast (ko04111; 89 genes; Cellular Processes; Cell growth and death), meiosis-yeast (ko04113; 79 genes; Cellular Processes; Cell growth and death).

***P. elaeocarpicola* genome has advantages in cellulose degradation**

The plant cell wall consists of structural proteins, aromatic compounds and carbohydrates such as pectin, cellulose, and hemicellulose, which acts as a front line against pathogenic fungi [44, 45]. CAZymes secreted by pathogenic fungi play a crucial role in plant cell wall degradation and are therefore important virulence factors [16]. CAZymes are composed of proteins with the ability to modify, degrade, or generate glycosidic bonds and are classified into six main classes: glycoside hydrolases



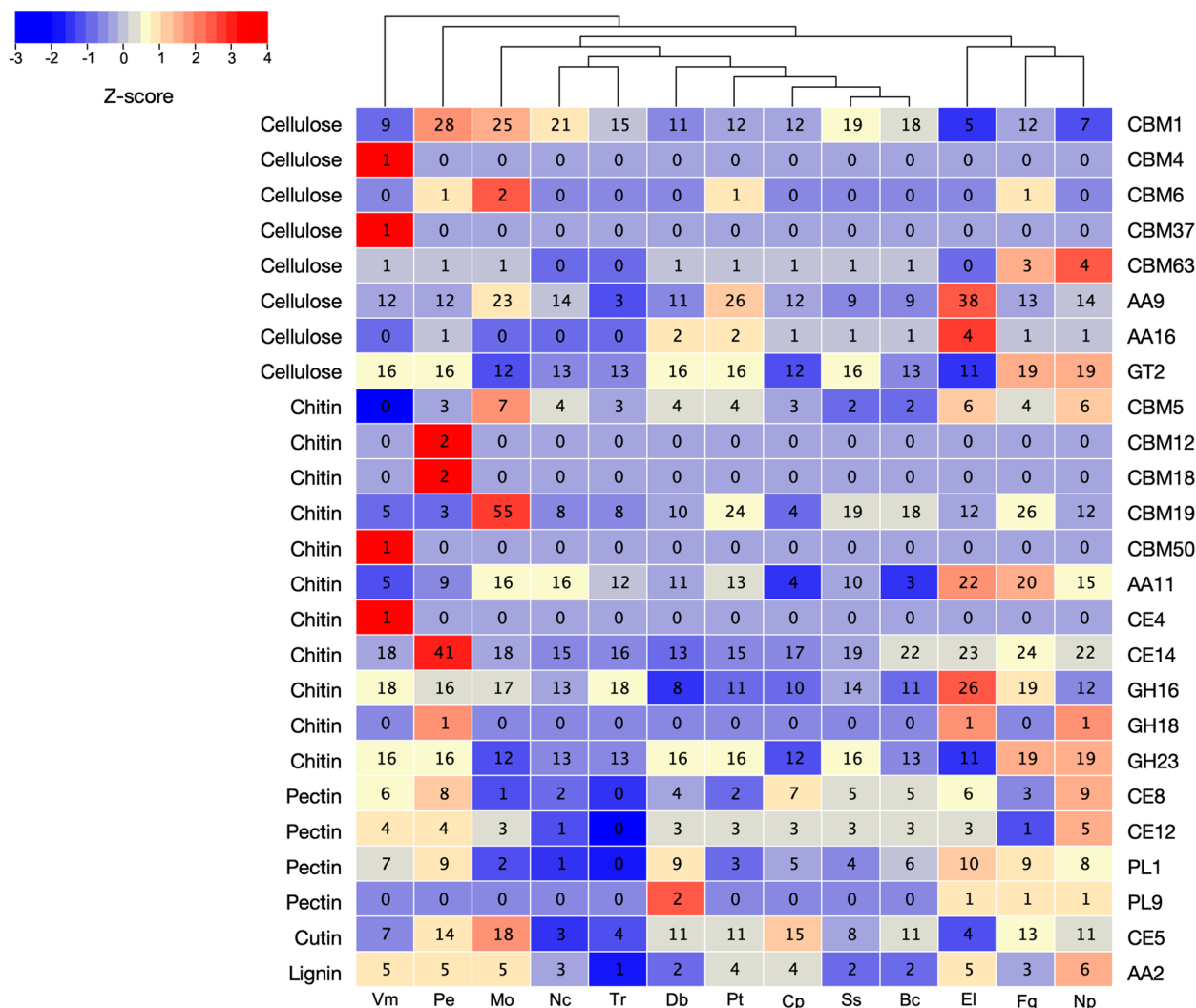
**Fig. 3** Phylogenetic relationship and orthologous gene cluster. A maximum likelihood phylogenetic tree was constructed based on 2925 single-copy ortholog genes showing the phylogenetic relationship of *P. elaeoarpicola* and 12 reference fungi. The colored bars and numbers represent orthologous proteins identified using Orthofinder for each fungal genome



**Fig. 4** Analysis of genome annotation and metabolic pathway of *P. elaeoarpicola*. **A** Gene Ontology (GO) annotation of *P. elaeoarpicola* genome; **B** KEGG pathway annotation of *P. elaeoarpicola* genome

(GHs), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), carbohydrate-binding modules (CBMs), and auxiliary activities (AAs) [11, 46]. Based on the importance of CAZymes, we predicted the CAZymes content in *P. elaeoarpicola* genome and compared it with those of 12 other representative fungi (Fig. 5, Table S1).

The CAZymes identification the existence of 213 different CAZyme families. Among them, the GH family (113 GH families) is the most abundant, followed by the GT family (41 GT families). In the GH family, the *P. elaeoarpicola* genome contained the highest number of GH5, GH43, and GH16 families, which were much higher than other tested fungi (Table S1). GH16 family is a large and



**Fig. 5** Hierarchical clustering of CAZyme families from *P. elaeocarpicola* and 12 other fungal genomes. Right side of heatmap showing the characterized subfamilies of CAZymes. GH, glycoside hydrolase; GT, glycosyl transferase; CE, carbohydrate esterase; AA, auxiliary activities; CBM, carbohydrate-binding module; PL, polysaccharide lyase. Fungal names (left to right): Vm, *Valsa mali*; Pe, *Pseudocryphonectria elaeocarpicola*; Mo, *Magnaporthe oryzae*; Nc, *Neurospora crassa*; Tr, *Trichoderma reesei*; Db, *Drepanopeziza brunnea*; Pt, *Pyrenophora teres f. teres*; Cp, *Cryphonectria parasitica*; Ss, *Sclerotinia sclerotiorum*; Bc, *Botrytis cinerea*; El, *Eutypa lata*; Fg, *Fusarium graminearum*; Np, *Neofusicoccum parvum*. Overrepresented (yellow to red) and underrepresented (yellow to blue) domains are depicted as Z-scores for each family. Data are from this study and from the carbohydrate-active enzyme database (<http://www.cazy.org/>)

taxonomically diverse family of glycosidases and trans-glycosidases that participate in various physiological processes, including in the degradation or remodeling of cell wall polysaccharides, nutrient absorption, invasion and pathogenesis, and regulation of immune responses. The subfamilies of GH16 in *P. elaeocarpicola* genome include GH16-1, GH16-3, GH16-4, GH16-10, GH16-18, GH16-19, GH16-20, and GH16-23, of which the genes coding for the subfamilies of GH16-1, GH16-18, and GH16-20 ranked the top three in terms of number. GH16-20 is a functional subfamily composed of plant

enzymes specific for xyloglucans [47]. The enzymes of this subfamily mainly consist of two main groups: xyloglucan endo-transglycosylases and xylan endohydrolases [48]. Xyloglucans are widely distributed in vascular plants and are the main component of the primary wall of all dicots and some monocots [48]. As shown in Fig. 5, *P. elaeocarpicola* contains large number of genes encoding cellulose degradation, with a total of 59 genes. Besides, GH5-7, which is important for hemicellulose degradation, was also identified (Table S1). These results indicate that *P. elaeocarpicola* is able to degrade cellulose,

hemicellulose, and xyloglucans in host plant cell walls effectively.

Hierarchical clustering analysis showed that *P. elaeocarpicola* has unique CAZymes profile (Fig. 5). Remarkably, *P. elaeocarpicola* genome contain the largest number of CBM1, CBM12, CBM18, and CE14 families among all tested fungi. CBM1 family is widely present in fungal cellulases and can promote cellulose degradation by increasing the local concentration of cellulase [49, 50]. The number of cutinases genes (CE5 family) in *P. elaeocarpicola* was similar to *C. parasitica*, which is consistent with the fact that pathogenic fungi of Cryptomeriaceae family penetrate wounds by degrading plant cuticles. In addition, compared with *C. parasitica*, *P. elaeocarpicola* contains comparable number of genes involved in pectin, cuticle and lignin degradation, but the number of genes encoding degradation of chitin and cellulose were much more. These results showed that *P. elaeocarpicola* may possess a greater potential to degrade chitin and cellulose in comparison with *C. parasitica*.

#### ***P. elaeocarpicola* genome is rich in potentially virulent related secret proteins**

It has been shown that fungi accomplish adaptation to the host environment, acquisition of nutrients and defense against immune responses through secreted proteins and other molecules. Especially, the secreted proteins produced by pathogenic fungi can facilitate colonization of host tissue and even kill host cells [40]. Therefore, the identification and functional annotation of pathogenic fungi secretome is important for identifying potential virulence factors. In this study, SIGNALP and TMHMM were used to predict secreted proteins in the genome of *P. elaeocarpicola*. There are 10,894 predicted genes encoding proteins in the *P. elaeocarpicola* genome with 841 putative secreted proteins, accounting for 7.72% of the genome. Functional enrichment showed that 367 secreted proteins could be assigned GO terms. Among them, proteins involved in hydrolase activity, hydrolyzing

O-glycosyl compounds, pectin metabolic process, carbohydrate catabolic process, and cell wall polysaccharide metabolic process are significantly overrepresented in the *P. elaeocarpicola* secretome (Fig. S1). In addition, functional analysis showed that, *P. elaeocarpicola* secretome contains homologs of known pathogenic effectors of other plant pathogens (Table 2), including Gell1 and MoPMO9A of *M. oryzae* [51, 52], Fgl1 and FgKP4L-2 of *Fusarium graminearum* [53, 54], and the pelB of *Colletotrichum gloeosporioides* and so on [55–58].

Based on secreted proteins, a total of 236 effector proteins in the *P. elaeocarpicola* genome were predicted by EffectorP. Of these 236 effector proteins, 169 proteins were annotated as apoplastic effector, 59 proteins were annotated as cytoplasmic effector, and eight proteins were annotated as apoplastic or cytoplasmic effector. Subsequently, we matched 236 effector proteins to the Non-Redundant Protein Database (NR), and 209 proteins were matched. Among them, 171 proteins were annotated to *C. parasitica*, of which 63 proteins were functionally known with the identity majority above 81.8% (Table S2). This result is consistent with the phylogenetic analysis of *P. elaeocarpicola*. Among these 171 effector proteins, a total of 34 proteins were annotated as GHs, encompassing 10 families, with GH28 (12 proteins) accounting for the largest proportion. GH28 family contains polygalacturonases (PG), rhamnogalacturonases (RG), and xylogalacturonases (XG), which have the function of degrading the pectin layer of plant cell walls. Many GH28 family members have been identified and characterized as virulence factors for pathogenic fungi. In addition, we have annotated four GH7 family proteins in the *P. elaeocarpicola* genome (Table S2).

Additionally, 27 effector proteins were specific in *P. elaeocarpicola*. Function prediction was performed for 27 proteins using the InterProScan tool, among which GME8245\_g (Tripeptidyl-peptidase I and related peptidases, IPR050819), GME97\_g (Cerato-ulmin hydrophobin family, IPR010636) and GME6290\_g (HeLo

**Table 2** Putative effectors showing homology to known fungal effectors

<i>P. elaeocarpicola</i>	Annotation	Domain	Reference
GME4321_g	family 72 glycoside hydrolase	Glucanoyltransferase (IPR004886)	Samalova, Marketa et al. (2017) [52]
GME4462_g	DnaJ-domain-containing protein	Chaperone DnaJ (IPR012724)	Wang, Jie et al. (2016) [57]
GME10643_g	glycosylhydrolase family 61-1	Auxiliary Activity family 9	Martinez-D'Alto, Alejandra et al. (2023) [51]
GME9214_g	family 5 carbohydrate esterase	Cutinase/acetylxyylan esterase (IPR000675)	Gong, Yingdi et a (2022) [55]
GME5618_g	alpha/beta-hydrolase	Alpha/Beta hydrolase fold (IPR029058)	Sella, Luca et al. (2014) [54]
GME5964_g	family 1 polysaccharide lyase	Pectin lyase family (IPR045032)	Yakoby, N et al. (2001) [58]
GME4212_g	killer toxin	Killer toxin Kp4/SMK (IPR011329)	Lu, Shunwen, and Justin D Faris (2019) [53]
GME4353_g	family 5 carbohydrate esterase	Cutinase/acetylxyylan esterase (IPR000675)	Lee, Miiin-Huey et al (2010) [56]



domain superfamily, IPR038305) were functionally annotated. Cerato-ulmin hydrophobin family may mediate contact and communication between fungi and environment [59]. In addition, this family includes cryparin, a cell-surface-associated hydrophobin secreted by filamentous fungi, trihydrophilin in the ergot *Claviceps fusiformis*, and cerato-ulmin, secretory toxin hydrophilin associated with Dutch elm disease [60–62]. These specific proteins most likely determine the specific function of *P. elaeocarpicola* and have high research value for further research on the evolution of *P. elaeocarpicola*. Taken together, these results suggest that *P. elaeocarpicola* would promote pathogenicity to host plants by secreting a large abundance of putative effector proteins.

**The secondary metabolism genes of *P. elaeocarpicola* are particularly abundant**

The secondary metabolites (SMs) secreted by fungi play a crucial role in nutrient acquisition, chemical warfare, communication, and ecological interactions. In phytopathogens, phytotoxic SMs are the key weapons during the invading of host plant, mainly including polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs) and terpenes. AntiSMASH7.1 was applied to

identify genes involved in secondary metabolite biosynthesis in the genome of *P. elaeocarpicola* (Table 3 and Table S3). The identification results showed that *P. elaeocarpicola* encodes a wide variety of secondary metabolism core enzymes, including 28 PKSs, 23 NRPSs and putative NRPS-like enzymes, seven terpenes, 23 Fungal-RiRP-likes, and five other enzymes. These core enzymes are distributed in 61 secondary metabolism gene clusters. A total of 44 transporter genes were identified in the 61 secondary metabolism gene clusters. As shown in Fig. 6A, we further compared the SMs genes of *P. elaeocarpicola* to 12 typical stem and foliar fungi with different lifestyles. Notably, *P. elaeocarpicola* genome contains abundant SMs genes among all tested fungi. These results indicated that *P. elaeocarpicola* have the great potential for SMs production.

The homology analysis of *P. elaeocarpicola* secondary metabolism gene clusters showed that out of 61 secondary metabolism gene clusters, a total of seven (more than 50% similarity) gene clusters were reported to be homologous. Among these seven clusters, cluster12 (GME1619\_g~GME1630\_g), cluster28 (GME6528\_g~GME6538\_g), and cluster52 (GME2317\_g~GME2323\_g) had alignment results with 100% similarity rate. Furthermore, these three secondary

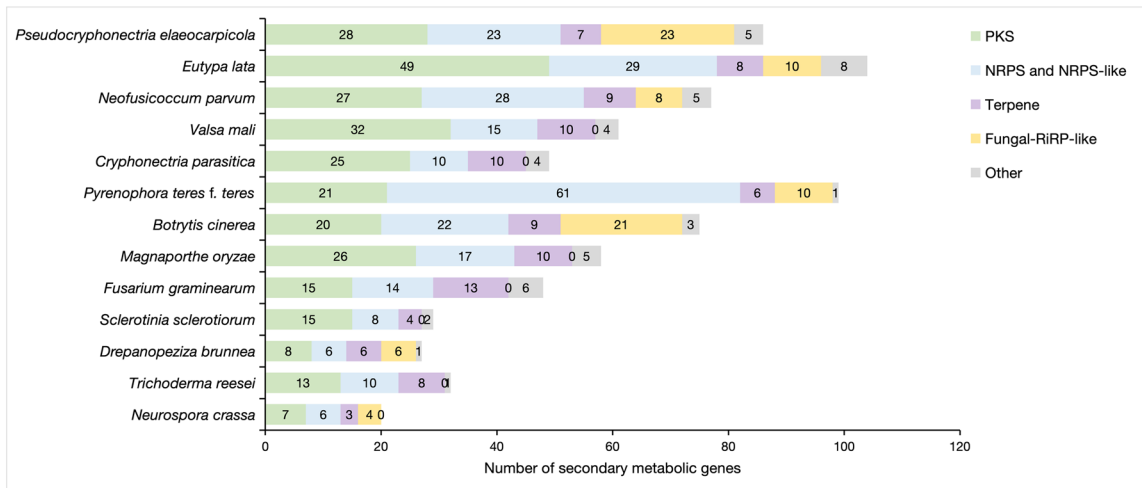
**Table 3** The core genes involved in the biosynthesis of secondary metabolites

Species	Scaffold	PKS	NRPS and NRPS-like	Terpene	Fungal-RiRP-like	Other	Total	
<i>P. elaeocarpicola</i>	Scaffold1	9	3	2	7	0		
	Scaffold2	6	3	1	3	3		
	Scaffold3	0	1	1	3	0		
	Scaffold4	3	3	1	0	2		
	Scaffold6	3	2	0	5	0		
	Scaffold7	1	1	0	1	0		
	Scaffold8	0	1	1	0	0		
	Scaffold9	3	4	0	1	0		
	Scaffold10	1	0	1	1	0		
	Scaffold11	1	2	0	1	0		
	Scaffold12	1	2	0	0	0		
	Scaffold13	0	0	0	1	0		
	Scaffold14	0	1	0	0	0		
	Total		28	23	7	23	5	86

(See figure on next page.)

**Fig. 6** The core secondary metabolism genes and pathogenic related homologous gene clusters in *P. elaeocarpicola*. **A** The core genes involved in the biosynthesis of secondary metabolites from *P. elaeocarpicola* and 12 other fungal genomes. NRPS and NRPS-like (Non-Ribosomal Peptide Synthase and NRPS-like fragment), PKS (Type I Polyketide Synthase and Type III Polyketide Synthase), Terpene, Fungal-RiPP-like (Fungal RiPP-likes), Other (indole, orf2\_PTase, phosphonate-like, phosphonate). Data are from this study and from the antiSMASH fungal v7.0 (<https://fungismash.secondarymetabolites.org/#!/start>); **B** The line with the arrow indicates the scheme of the cluster12 (GME1619\_g~GME1630\_g), cluster28 (GME6528\_g~GME6538\_g), and cluster52 (GME2317\_g~GME2323\_g) of *P. elaeocarpicola*. Domains were predicted using the InterProScan tool. Different color boxes indicate different domain categories. The red star represents the core gene. Bar = 2 kb

A



B

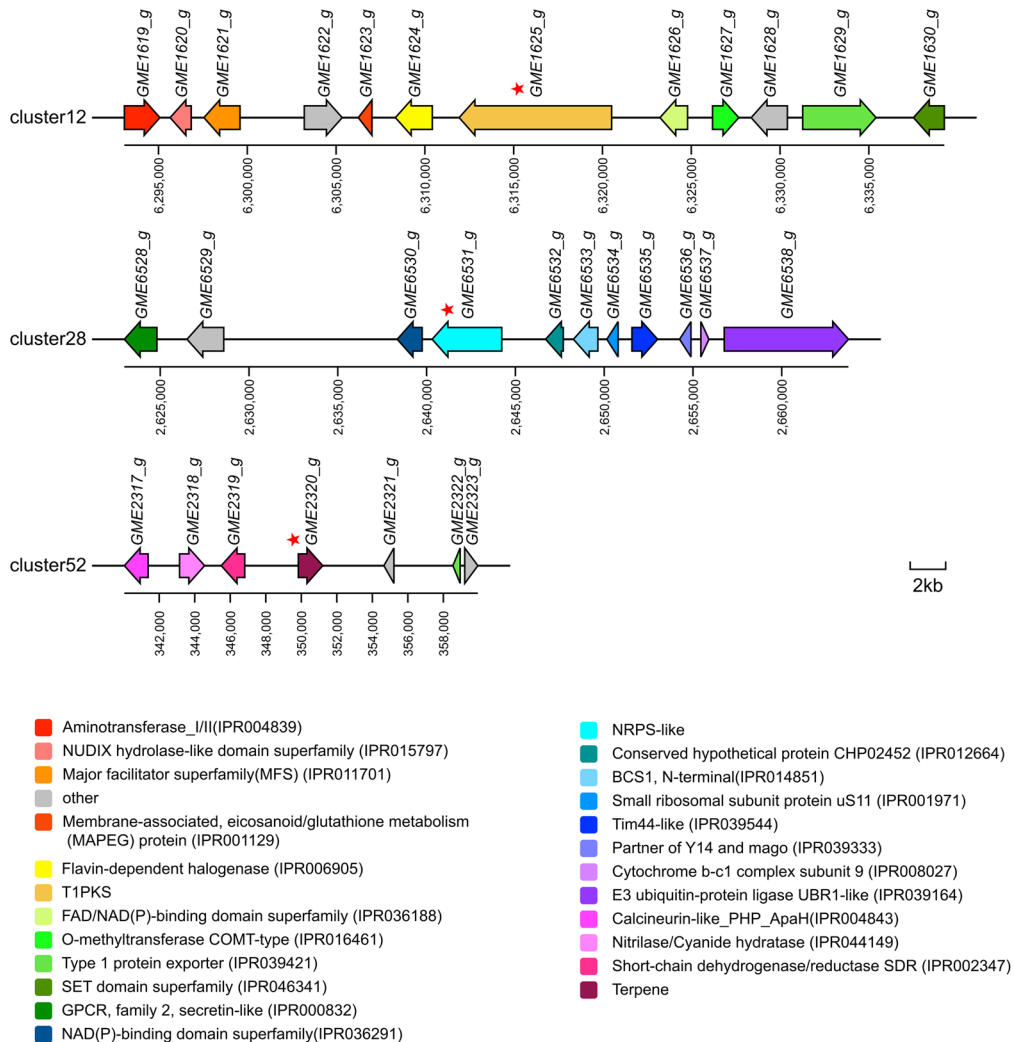


Fig. 6 (See legend on previous page.)

metabolism gene clusters were functional characterized (Fig. 6B). Cluster12 (T1PKS) was predicted to have 100% similar to the known ochratoxin A (OTA) biosynthesis gene cluster in *Aspergillus steynii*, which is one of the most important mycotoxins with strong toxicity and worldwide distribution [30]. Cluster52 (terpene) was predicted to be homologous to the known ACR toxin biosynthetic gene in *Alternaria alternata*, which was required for the production of ACR toxin and the pathogenicity of *A. alternata* [63]. Cluster28 (NRPS-like) was capable to synthesis of choline which affects mycelial morphology by controlling branch initiation and is an essential metabolite for the growth of filamentous fungi [64]. In summary, these results suggest that SMs genes play an essential role in the development and infection of *P. elaeocarpicola*.

Fungal cytochrome P450 (CYP) monooxygenase superfamily plays a crucial role in a wide range of process, such as primary metabolism, secondary metabolism, detoxification, and virulence [65]. In this study, a total of 152 putative CYPs were predicted in *P. elaeocarpicola* (Table S4). Compared with typical pathogenic fungi *C. parasitica* (118 CYPs), *Verticillium dahliae* (69 CYPs), and *M. oryzae* (131 CYPs), *P. elaeocarpicola* encodes a great number of CYPs. CYP gene localization analysis showed that 25 CYP genes distributed in 19 secondary metabolism gene clusters (Table S3). To further predict the potential of CYPs in the pathogenic process of *P. elaeocarpicola*, 152 CYP genes were searched in the PHI-base. The results showed that a total of 102 CYP genes were similar to known fungal pathogenic CYP genes. Of these, six genes are similar to the characterized fungal CYP genes, such as VdCYP1, which is participate in the in phytotoxin (compound sulfacetamide) biosynthesis in *V. dahliae* [66] (Table S5).

#### Determination of transporter proteins

Transporter proteins are a class of membrane proteins in which the ATP-binding cassette (ABC) and major facilitator superfamily (MFS) of transporters are common in filamentous fungi. Transporter proteins provide protection for filamentous fungi by resisting toxic compounds in the environment and transporting nutrients. In pathogenic fungi, membrane proteins can also mediate the secretion of host-specific toxins, against plant defensive compounds, and are important determinants in the pathogenic process [67]. Based on Transporter Collection Database [14], we identified 537 transporter genes in *P. elaeocarpicola* genome (Table S6).

MFS transporters are involved in the transport of a variety of substances such as organic and inorganic anions, monosaccharides and oligosaccharides, inositol, drugs, amino acids, organophosphorus esters and Krebs

cycle metabolites. *P. elaeocarpicola* genome has the highest number of MFS transporters, and more than 60 MFS transporters were predicted, including Drug: H<sup>+</sup> Antiporter-1 (DHA1), DHA2 and Sugar Porter (SP) family transporters (Table S7). Among them, SP family transporters are overrepresented in *P. elaeocarpicola* genome, with 23 genes, accounting for 37% of the MFS superfamily. This result suggested that *P. elaeocarpicola* might absorb more nutrients produced by host plants. Besides, the ability of DHA transporters to transport specific chemicals across the plasma membrane and to participate in the development of multidrug resistance suggest that *P. elaeocarpicola* can export more metabolites with high efficiency [68].

#### Transcription factors (TFs) in the analyzed genomes

Transcription factors (TFs) are sequence-specific DNA-binding proteins required for the regulation of gene expression and can regulate a variety of biological processes, such as maintaining cellular homeostasis in response to environmental changes, regulating growth and development, and interacting with other organisms [69, 70]. In pathogenic fungi, TFs act as upstream components that regulate various pathways associated with fungal virulence, including hyphal growth, sporulation, secondary metabolite biosynthesis, infection-related morphogenesis, melanization and so on [70]. In current study, *P. elaeocarpicola* encodes 417 putative TFs, including 191 Zn2Cys6-type TFs, 82 C2H2-type TFs, 21 bZIP-type TFs, five GATA-type TFs, five APSES (bHLH) -type TFs and so on (Table S8). These TFs families play important roles in fungal development and pathogenicity [70].

To further investigate the role of TFs on the pathogenicity of *P. elaeocarpicola*, a total of 304 fungal development and pathogenicity related TFs were searched in the PHI-base. Eventually, 284 TFs were found to have homologs in PHI-base of which 66 TFs showed high similarity to known pathogenic genes (Table S9). For example, two putative Zn2Cys6-type TFs, *GME10067\_g* and *GME6095\_g*, were significantly similar to *GPF1* (PHI:3307) and *CCA1* (PHI:3314) of *M. oryzae*, respectively. Knockout *GPF1* and *CCA1* were experimentally proven resulted in loss of pathogenicity in *M. oryzae* [71]. In addition, two putative bZIP-type TFs, *GME8542\_g* and *GME9227\_g*, were similar to virulence associated genes *TOXE* (PHI:233) in *Bipolaris zeicola* and MoAP1 (PHI:2142) in *M. oryzae*, respectively [72, 73]. The other virulence associated TFs including the putative APSES-type TFs *GME4946\_g* (Mstu1, PHI:2065, *M. oryzae*, loss of pathogenicity), the putative GATA-type TFs *GME766\_g* (*clnr1*, PHI:287, *Colletotrichum lindemuthianum*, loss of pathogenicity) and three putative

C2H2-type TFs *GME1376\_g* (Con7p, PHI:2039, *M. oryzae*, loss of pathogenicity), *GME1536\_g* (Vph1, PHI:7330, *V. dahliae*, loss of pathogenicity) and *GME7593\_g* (Rbf1, PHI:2237, *Ustilago maydis*, loss of pathogenicity) [74–78]. These results suggest that the TFs of *P. elaeocarpicola* have great potential to participate in regulation of the fungal virulence.

## Discussion

*P. elaeocarpicola* is one of the most destructive pathogenic fungi of *Elaeocarpus* spp. in Guangdong Province, China. Until now, no effective treatment or detailed pathogenic mechanisms has been established for this disease. We firstly sequenced the genome of *P. elaeocarpicola*. The high-quality genome sequence of *P. elaeocarpicola* is 44.49 Mb, which is similar to that of *C. parasitica* (43.9 Mb) in the same family. The genome analysis provided us with a repertoire of genes related to *P. elaeocarpicola* infection and colonization of the host, including genes involved in the degradation of the plant cell wall, biosynthesis of secondary metabolites, synthesis of secreted proteins and translocation factors. This study lay the foundation for future research on *P. elaeocarpicola* and provide important data for the design of efficient control strategies.

The CAZymes secreted by *P. elaeocarpicola* fraction include many enzymes involved in the degradation of key plant cell wall components. In addition to GH16, *P. elaeocarpicola* also showed a marked expansion in GH5, GH43, and GH28 families compared with other tested fungi. Currently, one of the largest cellulase families is GH5, with at least 56 subfamilies [79]. Because GH5 family was the first cellulase family to be described also known as “cellulase family A”. GH5 family is also the basis for hemicellulose degradation. *Heterobasidion irregulare* possesses two GH5 (b-mannanases), which play an important role in the degradation of softwood hemicellulose [80, 81]. GH43 family is also a large family in GH families, which is divided into 37 subfamilies. The xylanolytic enzyme of GH43 is involved in the breakdown of hemicellulose [82]. These results demonstrate the powerful ability of *P. elaeocarpicola* to degrade cellulose and hemicellulose in host plant cell walls. GH28 family is an important extracellular enzyme in pathogenic fungi that hydrolyzes the glycosidic bonds in pectin, the main structural component of plant cell wall [83]. Many GH28 family members have been found to play critical roles in virulence in a variety of pathogenic fungi: VdEPG1 in *V. dahliae* can modulate the biosynthesis of jasmonic acid (JA) by regulating *GhOPR9*, and act as a virulence factor to regulate the host immune response. GH28 family also play an important role during the infection of *Xanthomonas campestris* pv. *campestris*, *Alternaria citri* and

*B. cinerea* [84, 85]. In addition, as shown in Table S1, the number of genes for GH28 was generally higher in necrotrophic pathogens than in fungi of other lifestyles. This result may be caused by differences in fungal lifestyles. Necrotrophic pathogens vandalize cell wall components and degrade plant tissues to obtain carbon and energy. In contrast, biotrophic fungi will colonize intact plant cells while avoiding immune responses [86]. GH28, as the extracellular enzymes, trigger a series of plant immunological responses and stimulate host plants to produce hypersensitive immune responses or localized cell death mediated by polygalacturonase-inhibiting proteins (PGIPs) [83]. As a result, necrotrophic pathogens would appear to have a higher GH28 abundance than biotrophic fungi. In addition, previous studies have found that polygalacturonases of GH28 family contribute to the enhancement of virulence in *C. parasitica* [87]. Based on this result, it is highly likely that polygalacturonases belonging to the GH28 family in *P. elaeocarpicola* also contributes to virulence.

Previous studies on the interaction between pathogenic fungi of the Cryptomeriaceae family and host have found that the pathogenic fungi of this family have the ability to colonize through bark wounds [3]. In current study, CAZymes annotation results of *P. elaeocarpicola* are consistent with this feature of the family, and the significantly expand *P. elaeocarpicola* of the GH5, GH43, and GH28 family has great potential for colonization on host bark. Furthermore, based on the abundance of cellulose and hemicellulose degrading enzymes, *P. elaeocarpicola* has a higher capacity for plant cell wall degradation than *C. parasitica*, which is consistent with the fact that *P. elaeocarpicola* and *C. parasitica*, have different disease development time. Most diseased trees infected with *P. elaeocarpicola* died within five months, and smaller branches or twigs infected with *C. parasitica*, died within several months [3, 5].

Secreted proteases and membrane transporters play an important role in degradation of host structural proteins and nutrient absorption during plant infection, respectively. Our results revealed that *P. elaeocarpicola* genome contains numerous and functionally rich secreted proteins and effector proteins. Among the 171 effector proteins, we annotated four GH7 family proteins and 12 GH28 family proteins, of which GH28 family proteins has the highest number of proteins. Although the function of GH7 family proteins in plant pathogenic fungi is scarcely reported, the absence of PsGH7a has an indispensable influence on *Phytophthora sojae* virulence, which can lead to a reduction in the size of the lesion [88]. In addition, the *P. elaeocarpicola* secretome contains homologs of identified virulence secreted proteins of other plant pathogens as well as other pathogenesis-associated GH

family members, including Gel1 and MoPMO9A in *M. oryzae*, *SsCut1* in *S. sclerotiorum*, and so on (Table S2). Gel1 deficient mutant of *M. oryzae* could not produce conidia, and loss pathogenicity [52]. MoPMO9A, which is significantly expressed in the early stages of infection, promotes the binding of CBM to the fungal cell wall chitin, which in turn weakens the plant cell wall and facilitates pathogen invasion [51]. Besides, *GME9214\_g* in *P. elaeocarpicola* has a high homology with *SsCut1* in *S. sclerotiorum*. *SsCut1* gene enhances the pathogenicity of *S. sclerotiorum* through the enhancement of cutinase activity [55]. These predicted virulence-related proteins have a high likelihood of being involved in the pathogenic process of *P. elaeocarpicola*, and therefore can serve as target proteins to provide basic research materials for future studies of pathogenesis. The specific secreted protein of *P. elaeocarpicola*, *GME8245\_g*, is homologous to VmEP1 of *V. mali* in PHI-base. VmEP1 is able to manipulate plant immune mechanisms by inhibiting BAX-induced programmed cell death and targeting host pathogenesis-related protein 10 (PR10) [89]. Based on this result, *GME8245\_g* can be used as a target gene for interactions between pathogenic fungi and hosts for intensive study. Membrane transporters of pathogenic fungi could mitigate the damaging effects of defense compounds produced by host plants and transport nutrients. The *P. elaeocarpicola* genome contains more MFS superfamilies compared with ABC transporter families. Therefore, we speculate that the selective amplification of MFS superfamily and the loss of ABC transporters family in *P. elaeocarpicola* suggests that *P. elaeocarpicola* may possess multidrug resistance genes to combat specific antifungal compounds in ecological niche.

Secondary metabolites (SMs, also known as natural products), are produced by specific fungal taxa during secondary metabolism. SMs are important in fungal development and response to stressors (both abiotic and biotic) [29]. In phytopathogenic fungi, SMs can also enhance the infectiveness, and further improve the infection processes, such as Aflatoxin B, T-toxin, cercosporin toxin, fusaoctaxin A, and so on [32, 33, 90–92]. Our annotation shows that there are a large number of secondary metabolism gene clusters in *P. elaeocarpicola*, containing a total of 86 secondary metabolism backbone genes. Besides, cluster12 and cluster52 of *P. elaeocarpicola* were predicted to have 100% homology to the known toxin synthesis gene clusters ochratoxin A (OTA) and ACRTS2, respectively. OTA, mainly produced by *Aspergillus* and *Penicilium* species, is one of the most important mycotoxins that contaminates agricultural including cereal products, grapes and grapes products, cocoa, nuts, coffee, beverages, dried fruits, and cured meats [93]. OTA can cause many adverse effects on the health

of humans and livestock, such as strong nephrotoxicity, hepatotoxicity, teratogenicity, immunosuppression and so on [93–95]. As for ACR-toxin, it is a host-selective toxins (HSTs) produced by *Alternaria* species during infection, which causes leaf necrosis of rough lemon and is a decisive factor in pathogenicity [96]. Previous reports showed that *C. parasitica*, which is closely related to *P. elaeocarpicola*, could produce a range of compounds associated with virulence, including oxalic acid, tannases, laccases, cryparin and diaporthin [87]. In summary, functionally rich SMs are likely to play important roles in *P. elaeocarpicola* growth and development, pathogenicity, stress response and other cellular processes.

## Conclusions

Taken together, the present study is the first to sequence and report the genome of *P. elaeocarpicola*, the pathogenic fungus of *Elaeocarpus* spp. stem blight. Intensive analysis of this genome suggests that *P. elaeocarpicola* may employ a variety of virulence strategies, including the secretion of CAZymes, secreted proteases, SMs, and membrane transporters, during the process of colonization and infection. Both CAZymes and SMs genes were significantly expanded in *P. elaeocarpicola* compared to the majority of typical reference pathogenic fungi. The *P. elaeocarpicola* genome provides valuable basic research materials for the molecular pathogenesis study and control strategies of blight pathogens.

## Materials and methods

### Biological material, genomic DNA extraction

*Pseudocryphonectria elaeocarpicola* was originally isolated from diseased *Elaeocarpus apiculatus* barks (collection of *Elaeocarpus apiculatus* barks complied with local guidelines). The strain were deposited in China Forestry Culture Collection Center (CFCC, <http://cfcc.caf.ac.cn/>), and the specimens in the herbarium of the Chinese Academy of Forestry (CAF, <http://museum.caf.ac.cn/>). Genomic DNA was extracted by the protocol of Cetyltrimethylammonium Bromide (CTAB). The integrity, purity, and concentration of genomic DNA were measured using Qubit fluorometer, and agarose gel electrophoresis, respectively.

### Genome sequencing, assembly and component prediction

Genomic DNA was extracted and randomly interrupted, DNA fragments of the desired length were recovered by electrophoresis, and splices were added for cluster preparation, and finally sequenced on the machine. Subsequently, the DNBSEQ platform sequentially removes 40% of the reads with consecutive bases  $\leq 20$  in the original sequencing data, reads with a certain proportion of Ns (10% as default), adapter contamination, and duplication

contamination, and finally obtains Clean Reads. The raw data from PacBio sequencing are Polymerase Reads, which are filtered out to obtain usable Subreads and saved in bam format, such as sequencing splices and low-quality data. Consistent sequence analysis is performed on each ZMW to obtain more accurate data, i.e., circular consensus sequencing data (CCS, also known as insertion reads), which will be used directly for assembly, alignment, species classification, etc. Subreads less than 1000 bp in length are removed [97–99]. Sequencing data were assembled using a variety of software and can be divided into four sections: a. Subreads correct (Canu v1.5, FalconConsensus, SMRT Analysis v2.3.0); b. Corrected Reads Assembly (Canu v1.5, Falcon v0.3.0); c. Correct single base (GATK v1.6-13); d. Link Contig to Scaffold and fill gap (SSPACE\_Basic v2.0), resulting in highly confident assembled sequences [98, 100–102].

The genetic components of *P. elaeocarpicola* were predicted using the software Augustus v3.2.1 and GeneMarkes v4.21 [103, 104]. The rRNA was predicted by RNAmmer v1.2; the area of tRNA and its secondary structure were predicted by tRNAscan v1.3.1; and the sRNAs were obtained by Infernal software, which was compared with the Rfam database v9.1 [105–107]. Transposon sequences were predicted by RepeatMasker v4-0-6, RepeatProteinMasker, and buildXDFDatabase (a denovo method). Tandem Repeat sequences (TR) were predicted by Tandem Repeat Finder v4.04 [108].

#### Annotation of specific gene categories

In the current study, GO (2019-07-01) and KEGG (101) were used to analyze the GO category and pathway [109–111]. To identify secreted proteins of fungal, protein sequence of *P. elaeocarpicola* were subjected to the SignalP 4.1 and TMHMM 2.0. Based on these prediction results, EffectorP 3.0 was further used to predict effector proteins [112]. Transport proteins of fungal species in this study were predicted using TransportDB [14]. The transcription factors of *P. elaeocarpicola* were annotated based on IPRSCAN annotation combined with the collection of transcription factor-related IPR Numbers from Fungal Transcription Factor Database (FTFD). Align the putative CYP genes and transcription factors of *P. elaeocarpicola*, with the pathogen-host interaction database (PHI-base, <http://www.phi-base.org/>) to screen the potential fungal development and pathogenic related genes [113]. Gene density, GCskew, GCratio analysis, synteny analysis, and circle plot visualization were completed by TTools [114]. The amino acid sequences of genes were compared with the carbohydrate-active enzyme database to identify genes encoding putative carbohydrate-active enzymes [46].

#### Gene family analysis

The reference species genomic data used in this study were downloaded from the NCBI website (<https://www.ncbi.nlm.nih.gov/genome/>) (Table S10). Orthofinder v2.5.5 was used to identify the single-copy orthogroup within *P. elaeocarpicola* and all reference fungi [115]. Multiple sequence alignments of each single-copy orthogroup were performed using MAFFT v7.520 [116]. The maximum-likelihood (ML) tree was inferred with FastTree v2.1.11 [117]. The obtained phylogenomic tree was visualized with FigTree v1.4.4.

#### Secondary metabolism genes and clusters analysis

The secondary metabolism gene clusters of *P. elaeocarpicola* were predicted using the antiSMASH fungal v7.0 [10]. The CYPs genes were predicted using Cytochrome P450 Engineering Database [12]. The domain structures of cluster12, cluster28, and cluster52 were annotated using the InterProScan tool (<https://www.ebi.ac.uk/interpro/>).

#### Accession codes

The complete genome of *Pseudocryphonectria elaeocarpicola* is available at CNCB (China National Center for Bioinformatics <https://www.cncb.ac.cn>) with accession numbers: GWHESHB00000000.

#### Abbreviations

CAZymes	Carbohydrate-active enzymes
SM	Secondary metabolite
NRPS	Non-ribosomal peptide synthetase genes
PKS	Polyketide synthase
CYPs	Cytochrome P450 related enzyme
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes
PHI-base	Pathogen-host interactions database
MFS	Major facilitator superfamily
ABC	ATP-binding cassette
GHs	Glycoside hydrolases
GTs	Glycosyltransferases
PLs	Polysaccharide lyases
CEs	Carbohydrate esterases
CBMs	Carbohydrate-binding modules
AAs	Auxiliary activities
NR	Non-Redundant Protein Database
PG	Polygalacturonases
RG	Rhamnogalacturonases
XG	Xylogalacturonases
OTA	Ochratoxin A
DHA	Drug:H <sup>+</sup> Antiporter-1
SP	Sugar Porter
TFs	Transcription factors
HSTs	Host-selective toxins
PGIPs	Polygalacturonase-inhibiting proteins
CTAB	Cetyltrimethylammonium Bromide

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-024-10615-5>.

Additional file 1: Figure S1. Gene Ontology (GO) terms of secretory proteins in *P. elaeoarpicola* genome.

Additional file 2: Table S1. Number of carbohydrate-active enzyme modules of *P. elaeoarpicola* and 12 other fungi according to the CAZY database.

Additional file 3: Table S2. Effector proteins annotated in the NR database.

Additional file 4: Table S3. Inventory of secondary metabolism gene clusters in *P. elaeoarpicola*.

Additional file 5: Table S4. Inventory of Cytochrome P450 related enzyme (CYPs) in *P. elaeoarpicola*.

Additional file 6: Table S5. Putative CYP genes showing homology to known fungal CYP genes.

Additional file 7: Table S6. Distribution of genes coding for membrane transporter families in *P. elaeoarpicola*.

Additional file 8: Table S7. Distribution of genes coding for family of the Major Facilitator Superfamily (MFS) in *P. elaeoarpicola*.

Additional file 9: Table S8. Inventory of Transcription Factors (TFs) in *P. elaeoarpicola*.

Additional file 10: Table S9. Predicted results of pathogenesis-related transcription factors (TFs) in the PHI-base.

Additional file 11: Table S10. Reference fungi information.

### Acknowledgements

We thank all participants in this study for their invaluable contribution.

### Authors' contributions

DX, HH, DZ and CT designed the experiment, provided experimental materials and were responsible for project management and financial support; HH provided experimental materials; YY was responsible for data management and analysis, the production of tables and figures, and wrote the initial manuscript; YY, DX and CT were responsible for the revision of the manuscript. All authors read and approved the final manuscript.

### Funding

This research was funded by Forestry Science and Technology Innovation Project of Guangdong (2020KJX004) and the Guangdong Basic and Applied Basic Research Foundation (2023A1515011968).

### Availability of data and materials

The complete genome of *Pseudocryphonectria elaeoarpicola* is available at CNCB (China National Center for Bioinformation <https://www.cncb.ac.cn>) with accession numbers: GWHESH00000000.

### Declarations

#### Ethics approval and consent to participate

We declare that the collection of the *Elaeocarpus* spp. barks complied with the guidelines of Guangdong Province, China.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>State Key Laboratory of Efficient Production of Forest Resources, Beijing Forestry University, Beijing 100083, China. <sup>2</sup>Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou 510520, Guangdong, China.

Received: 4 May 2024 Accepted: 12 July 2024

Published online: 24 July 2024

### References

- Chen SF, Wingfield MJ, Li GQ, Liu FF. *Corticomorbus Sinomyrti* gen. et sp. nov. (Cryphonectriaceae) pathogenic to native *Rhodomyrtus tomentosa* (Myrtaceae) in South China. *Plant Pathol.* 2016;65:1254–66.
- Gryzenhout M, Myburg H, Wingfield BD, Wingfield MJ. Cryphonectriaceae (Diaporthales), a new family including *Cryphonectria*, *Chrysoporthe*, *Endothia* and allied genera. *Mycologia.* 2006;98(2):239–49.
- Rigling D, Prospero S. *Cryphonectria parasitica*, the causal agent of chestnut blight: invasion history, population biology and disease control. *Mol Plant Pathol.* 2018;19:7–20.
- Wang W, Li GQ, Liu QL, Chen SF. Cryphonectriaceae on Myrtales in China: phylogeny, host range, and pathogenicity. *Persoonia.* 2020;45:101–31.
- Huang H-Y, Huang H-H, Zhao D-Y, Shan T-J, Hu L-L. *Pseudocryphonectria Elaeoarpicola* gen. et sp. nov. (Cryphonectriaceae, Diaporthales) causing stem blight of *Elaeocarpus* spp. in China. *MC.* 2022;91:67–84.
- Cornell MJ, Alam I, Soanes DM, Wong HM, Hedeler C, Paton NW, et al. Comparative genome analysis across a kingdom of eukaryotic organisms: specialization and diversification in the fungi. *Genome Res.* 2007;17:1809–22.
- Galagan JE, Henn MR, Ma L-J, Cuomo CA, Birren B. Genomics of the fungal kingdom: insights into eukaryotic biology. *Genome Res.* 2005;15:1620–31.
- Sharma KK. Fungal genome sequencing: basic biology to biotechnology. *Crit Rev Biotechnol.* 2016;36:743–59.
- Haridas S, Albert R, Binder M, Bloem J, LaButti K, Salamov A, et al. 101 Dothideomycetes genomes: a test case for predicting lifestyles and emergence of pathogens. *Stud Mycol.* 2020;96:141–53.
- Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, et al. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* 2023;51:W46–50.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* 2009;37(Database):D233–8.
- Fischer M, Knoll M, Sirim D, Wagner F, Funke S, Pleiss J. The cytochrome P450 engineering database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics.* 2007;23:2015–7.
- Petersen TN, Brunak S, Von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011;8:785–6.
- Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res.* 2021;49:D461–7.
- Doehlemann G, Ökmen B, Zhu W, Sharon A. Plant pathogenic fungi. *Microbiol Spectr.* 2017;5:5.1.14.
- Douaiher M-N, Nowak E, Durand R, Halama P, Reignault P. Correlative analysis of *Mycosphaerella Graminicola* pathogenicity and cell wall-degrading enzymes produced *in vitro*: the importance of xylanase and polygalacturonase. *Plant Pathol.* 2007;56:79–86.
- Kikot GE, Hours RA, Alconada TM. Contribution of cell wall degrading enzymes to pathogenesis of *Fusarium graminearum*: a review. *J Basic Microbiol.* 2009;49:231–41.
- Quoc NB, Bao Chau NN. The role of cell wall degrading enzymes in pathogenesis of *Magnaporthe oryzae*. *CPPS.* 2017;18:18.
- Bruno G, Sparapano L. Effects of three esca-associated fungi on *Vitis vinifera* L.: III. Enzymes produced by the pathogens and their role in fungus-to-plant or in fungus-to-fungus interactions. *Physiol Mol Plant Pathol.* 2006;69:182–94.
- Gibson DM, King BC, Hayes ML, Bergstrom GC. Plant pathogens as a source of diverse enzymes for lignocellulose digestion. *Curr Opin Microbiol.* 2011;14:264–70.
- Zhang H, Wu Q, Cao S, Zhao T, Chen L, Zhuang P, et al. A novel protein elicitor (SsCut) from *Sclerotinia Sclerotiorum* induces multiple defense responses in plants. *Plant Mol Biol.* 2014;86:495–511.

22. Brito N, Espino JJ, González C. The Endo- $\beta$ -1,4-Xylanase Xyn11A is required for virulence in *Botrytis Cinerea*. *MPMI*. 2006;19:25–32.
23. Fernández-Acero FJ, Colby T, Harzen A, Carbú M, Wieneke U, Cantoral JM, et al. 2-DE proteomic approach to the *Botrytis cinerea* secretome induced with different carbon sources and plant-based elicitors. *Proteomics*. 2010;10:2270–80.
24. Kema GHJ, Van Der Lee TAJ, Mendes O, Verstappen ECP, Sandbrink H, Van Der Burgt A, et al. Large-scale gene discovery in the Septoria Tritici Blotch Fungus *Mycosphaerella graminicola* with a focus on in Planta expression. *MPMI*. 2008;21:1249–60.
25. Chen L-Q, Hou B-H, Lalonde S, Takanaga H, Hartung ML, Qu X-Q, et al. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*. 2010;468:527–32.
26. Fan J, Urban M, Parker JE, Brewer HC, Kelly SL, Hammond-Kosack KE, et al. Characterization of the sterol 14 $\alpha$ -demethylases of *Fusarium Graminearum* identifies a novel genus-specific CYP 51 function. *New Phytol*. 2013;198:821–35.
27. Morales-Cruz A, Amrine KCH, Blanco-Ulate B, Lawrence DP, Travadon R, Rolshausen PE, et al. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genomics*. 2015;16:469.
28. Yin Z, Liu H, Li Z, Ke X, Dou D, Gao X, et al. Genome sequence of *Valsa* canker pathogens uncovers a potential adaptation of colonization of woody bark. *New Phytol*. 2015;208:1202–16.
29. Keller NP. Fungal secondary metabolism: regulation, function and drug discovery. *Nat Rev Microbiol*. 2019;17:167–80.
30. Gil-Serna J, García-Díaz M, Vázquez C, Patiño B. Description of an orthologous cluster of ochratoxin biosynthetic genes in *Aspergillus* and *Penicillium* species. A comparative analysis. *Int J Food Microbiol*. 2018;268:35–43.
31. Sumarah MW. The deoxynivalenol challenge. *J Agric Food Chem*. 2022;70:9619–24.
32. Wu D, Oide S, Zhang N, Choi MY, Turgeon BG. ChLae1 and ChVel1 regulate T-toxin production, virulence, oxidative stress response, and development of the Maize Pathogen *Cochliobolus heterostrophus*. *PLoS Pathog*. 2012;8:e1002542.
33. Yu J, Chang P-K, Ehrlich KC, Cary JW, Bhatnagar D, Cleveland TE, et al. Clustered pathway genes in aflatoxin biosynthesis. *Appl Environ Microbiol*. 2004;70:1253–62.
34. Maguvu TE, Travadon R, Cantu D, Trouillas FP. Whole genome sequencing and analysis of multiple isolates of *Ceratocystis destructans*, the causal agent of *Ceratocystis* canker of almond in California. *Sci Rep*. 2023;13:14873.
35. Siewiers V, Viaud M, Jimenez-Teja D, Collado IG, Gronover CS, Pradier J-M, et al. Functional analysis of the cytochrome P450 monooxygenase gene *bcbot1* of *Botrytis Cinerea* indicates that Botrydial is a strain-specific virulence factor. *MPMI*. 2005;18:602–12.
36. Zhang X, Guo J, Cheng F, Li S. Cytochrome P450 enzymes in fungal natural product biosynthesis. *Nat Prod Rep*. 2021;38:1072–99.
37. Del Sorbo G, Schoonbeek H, De Waard MA. Fungal transporters involved in Efflux of natural toxic compounds and fungicides. *Fungal Genet Biol*. 2000;30:1–15.
38. Stefanato FL, Abou-Mansour E, Buchala A, Kretschmer M, Mosbach A, Hahn M, et al. The ABC transporter BcatrB from *Botrytis Cinerea* exports camalexin and is a virulence factor on *Arabidopsis thaliana*. *Plant J*. 2009;58:499–510.
39. Vela-Corcía D, Aditya Srivastava D, Dafa-Berger A, Rotem N, Barda O, Levy M. MFS transporter from *Botrytis cinerea* provides tolerance to glucosinolate-breakdown products and is required for pathogenicity. *Nat Commun*. 2019;10:2886.
40. McCotter SW, Horianopoulos LC, Kronstad JW. Regulation of the fungal secretome. *Curr Genet*. 2016;62:533–45.
41. Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol*. 2015;66:513–45.
42. Tan K-C, Oliver RP, Solomon PS, Moffat CS. Proteinaceous necrotrophic effectors in fungal virulence. *Funct Plant Biol*. 2010;37:907.
43. Crouch JA, Dawe A, Aerts A, Barry K, Churchill ACL, Grimwood J, et al. Genome sequence of the chestnut blight fungus *Cryphonectria parasitica* EP155: a fundamental resource for an archetypical invasive plant pathogen. *Phytopathology*<sup>®</sup>. 2020;110:1180–8.
44. Caffall KH, Mohnen D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res*. 2009;344:1879–900.
45. Höfte H, Voxeur A. Plant cell walls. *Curr Biol*. 2017;27:R865–70.
46. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res*. 2022;50:D571–7.
47. Eklöf JM, Brumer H. The *XTH* Gene Family: an update on enzyme structure, function, and phylogeny in Xyloglucan Remodeling. *Plant Physiol*. 2010;153:456–66.
48. Baumann MJ, Eklöf JM, Michel G, Kallas ÅM, Teeri TT, Czjzek M, et al. Structural evidence for the evolution of xyloglucanase activity from Xyloglucan *Endo*-transglycosylases: Biological implications for cell wall metabolism. *Plant Cell*. 2007;19:1947–63.
49. Oliveira C, Carvalho Y, Domingues L, Gama FM. Recombinant CBM-fusion technology — applications overview. *Biotechnol Adv*. 2015;33:358–69.
50. Jia H, Feng X, Huang J, Guo Y, Zhang D, Li X, et al. Recombinant family 1 carbohydrate-binding modules derived from fungal cellulase enhance enzymatic degradation of lignocellulose as novel effective accessory protein. *Front Microbiol*. 2022;13:876466.
51. Martinez-D'Alto A, Yan X, Detomasi TC, Saylor RI, Thomas WC, Talbot NJ, et al. Characterization of a unique polysaccharide monooxygenase from the plant pathogen *Magnaporthe oryzae*. *Proc Natl Acad Sci USA*. 2023;120:e2215426120. <https://doi.org/10.1073/pnas.2215426120>.
52. Samalova M, Mérida H, Vilaplana F, Bulone V, Soanes DM, Talbot NJ, et al. The  $\beta$ -1,3-glucanosyltransferases (Gels) affect the structure of the rice blast fungal cell wall during appressorium-mediated plant infection. *Cell Microbiol*. 2017;19:e12659. <https://doi.org/10.1111/cmi.12659>.
53. Lu S, Faris JD. *Fusarium Graminearum* KP4-like proteins possess root growth-inhibiting activity against wheat and potentially contribute to fungal virulence in seedling rot. *Fungal Genet Biol*. 2019;123:1–13. <https://doi.org/10.1016/j.fgb.2018.11.002>.
54. Sella L, Gazzetti K, Castiglioni C, Schäfer W, Favaron F. *Fusarium Graminearum* possesses virulence factors common to fusarium head blight of wheat and seedling rot of soybean but differing in their impact on disease severity. *Phytopathology*<sup>®</sup>. 2014;104:1201–7. <https://doi.org/10.1094/PHYTO-12-13-0355-R>.
55. Gong Y, Fu Y, Xie J, Li B, Chen T, Lin Y, et al. *Sclerotinia Sclerotiorum* *SsCut1* modulates virulence and cutinase activity. *J Fungi*. 2022;8:526. <https://doi.org/10.3390/jof8050526>.
56. Lee M-H, Chiu C-M, Roubtsova T, Chou C-M, Bostock RM. Overexpression of a redox-regulated cutinase gene, *MiCUT1*, increases virulence of the Brown Rot Pathogen *Monilinia fructicola* on *Prunus* spp. *MPMI*. 2010;23:176–86. <https://doi.org/10.1094/MPMI-23-2-0176>.
57. Wang J, Ying S, Hu Y, Feng M. Mas5, a homologue of bacterial DnaJ, is indispensable for the host infection and environmental adaptation of a filamentous fungal insect pathogen. *Environ Microbiol*. 2016;18:1037–47. <https://doi.org/10.1111/1462-2920.13197>.
58. Yakoby N, Beno-Moualem D, Keen NT, Dinoor A, Pines O, Prusky D. *Colletotrichum gloeosporioides pelB* is an important virulence factor in avocado fruit-fungus interaction. *MPMI*. 2001;14:988–95. <https://doi.org/10.1094/MPMI.2001.14.8.988>.
59. Whiteford JR, Spanu PD. The Hydrophobin HCF-1 of *Cladosporium Fulvum* is required for efficient water-mediated dispersal of Conidia. *Fungal Genet Biol*. 2001;32:159–68.
60. McCabe PM, Van Alfen NK. Secretion of Cryparin, a fungal Hydrophobin. *Appl Environ Microbiol*. 1999;65:5431–5.
61. Temple B, Horgen PA, Bernier L, Hintz WE. Cerato-ulmin, a hydrophobin secreted by the causal agents of Dutch Elm Disease, is a parasitic fitness factor. *Fungal Genet Biol*. 1997;22:39–53.
62. De Vries OMH, Moore S, Arntz C, Wessels JGH, Tudzynski P. Identification and characterization of a tri-partite hydrophobin from *Claviceps fusiformis*: a novel type of class II hydrophobin. *Eur J Biochem*. 1999;262:377–85.
63. Izumi Y, Ohtani K, Miyamoto Y, Masunaka A, Fukumoto T, Gomi K, et al. A Polyketide synthase gene, *ACRTS2*, is responsible for biosynthesis of host-selective ACR-toxin in the rough lemon pathotype of *Alternaria alternata*. *MPMI*. 2012;25:1419–29.



64. Markham P, Robson GD, Bainbridge BW, Trinci APJ. Choline: Its role in the growth of filamentous fungi and the regulation of mycelial morphology. *FEMS Microbiol Letters*. 1993;104:287–300.
65. Veith A, Moorthy B. Role of cytochrome P450s in the generation and metabolism of reactive oxygen species. *Curr Opin Toxicol*. 2018;7:44–51.
66. Zhang D-D, Wang X-Y, Chen J-Y, Kong Z-Q, Gui Y-J, Li N-Y, et al. Identification and characterization of a pathogenicity-related gene VdCYP1 from *Verticillium Dahliae*. *Sci Rep*. 2016;6:27979.
67. Stergiopoulos I, Zwiers L-H, De Waard MA. Secretion of natural and synthetic toxic compounds from filamentous fungi by membrane transporters of the ATP-binding cassette and major facilitator superfamily. In: Logrieco A, Bailey JA, Corazza L, Cooke BM, editors. *Mycotoxins in plant disease*. Dordrecht: Springer Netherlands; 2002. p. 719–34.
68. Sá-Correia I, Dos Santos SC, Teixeira MC, Cabrito TR, Mira NP. Drug:H<sup>+</sup> antiporters in chemical stress response in yeast. *Trends Microbiol*. 2009;17:22–31.
69. Charoensawan V, Wilson D, Teichmann SA. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res*. 2010;38:7364–77.
70. John E, Singh KB, Oliver RP, Tan K. Transcription factor control of virulence in phytopathogenic fungi. *Mol Plant Pathol*. 2021;22:858–81.
71. Lu J, Cao H, Zhang L, Huang P, Lin F. Systematic analysis of Zn2Cys6 transcription factors required for development and pathogenicity by high-throughput gene knockout in the rice blast fungus. *PLoS Pathog*. 2014;10:e1004432.
72. Guo M, Chen Y, Du Y, Dong Y, Guo W, Zhai S, et al. The bZIP transcription factor MoAP1 mediates the oxidative stress response and is critical for pathogenicity of the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathog*. 2011;7:e1001302.
73. Pedley KF, Walton JD. Regulation of cyclic peptide biosynthesis in a plant pathogenic fungus by a novel transcription factor. *Proc Natl Acad Sci USA*. 2001;98:14174–9.
74. Heimel K, Scherer M, Vranes M, Wahl R, Pothirata C, Schuler D, et al. The transcription factor Rbf1 is the master regulator for b-mating type controlled pathogenic development in *Ustilago maydis*. *PLoS Pathog*. 2010;6:e1001035.
75. Nishimura M, Fukada J, Moriwaki A, Fujikawa T, Ohashi M, Hibi T, et al. Mst1, an APSES transcription factor, is required for appressorium-mediated infection in *Magnaporthe Grisea*. *Biosci Biotechnol Biochem*. 2009;73:1779–86.
76. Odenbach D, Breth B, Thines E, Weber RWS, Anke H, Foster AJ. The transcription factor Con7p is a central regulator of infection-related morphogenesis in the rice blast fungus *Magnaporthe Grisea*. *Mol Microbiol*. 2007;64:293–307.
77. Pellier A, Laugé R, Veneault-Fourrey C, Langin T. CLNR1, the AREA/NIT2-like global nitrogen regulator of the plant fungal pathogen *Colletotrichum lindemuthianum* is required for the infection cycle. *Mol Microbiol*. 2003;48:639–55.
78. Sarmiento-Villamil JL, Prieto P, Klosterman SJ, García-Pedrajas MD. Characterization of two homeodomain transcription factors with critical but distinct roles in virulence in the vascular pathogen *Verticillium Dahliae*. *Mol Plant Pathol*. 2018;19:986–1004.
79. Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B. Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol*. 2012;12:186.
80. Butterfield B. The structure of wood: form and function. In: *Primary Wood Processing*. Springer, Dordrecht ; 2006. p. 1–22.
81. Olson Å, Aerts A, Asiegbu F, Belbahri L, Bouzid O, Broberg A, et al. Insight into trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New Phytol*. 2012;194:1001–13.
82. Morais MAB, Coines J, Domingues MN, Pirolla RAS, Tonoli CCC, Santos CR, et al. Two distinct catalytic pathways for GH43 xylanolytic enzymes unveiled by X-ray and QM/MM simulations. *Nat Commun*. 2021;12:367.
83. Sprockett DD, Piontkivska H, Blackwood CB. Evolutionary analysis of glycosyl hydrolase family 28 (GH28) suggests lineage-specific expansions in necrotrophic fungal pathogens. *Gene*. 2011;479:29–36.
84. Wang L, Rong W, He C. Two *Xanthomonas* Extracellular polygalacturonases, PghAxc and PghBxc, are regulated by type III secretion regulators HrpX and HrpG and are required for virulence. *MPMI*. 2008;21:555–63.
85. Liu S, Liu R, Lv J, Feng Z, Wei F, Zhao L, et al. The glycoside hydrolase 28 member VdEPG1 is a virulence factor of *Verticillium Dahliae* and interacts with the jasmonic acid pathway-related gene *GhOPR9*. *Mol Plant Pathol*. 2023;24:1238–55.
86. Oliver RP, Ipcho SVS. *Arabidopsis* pathology breathes new life into the necrotrophs-vs.-biotrophs classification of fungal pathogens. *Mol Plant Pathol*. 2004;5:347–52.
87. Lovat C, Donnelly DJ. Mechanisms and metabolomics of the host–pathogen interactions between Chestnut (*Castanea* species) and Chestnut blight (*Cryphonectria parasitica*). *Forest Pathol*. 2019;49:e12562.
88. Tan X, Hu Y, Jia Y, Hou X, Xu Q, Han C, et al. A conserved glycoside hydrolase family 7 cellobiohydrolase PsGH7a of *Phytophthora sojae* is required for full virulence on soybean. *Front Microbiol*. 2020;11:1285.
89. Wang W, Wang S, Gong W, Lv L, Xu L, Nie J, et al. *Valsa Mali* secretes an effector protein VmEP1 to target a K homology domain-containing protein for virulence in apple. *Mol Plant Pathol*. 2022;23:1577–91.
90. Ayada H, Dhoui B, Mazouz H, El Harrak A, Jaiti F, Ouhmidou B, et al. In silico comparative genomic analysis unravels a new candidate protein arsenal specifically associated with *Fusarium oxysporum* f. sp. *abedinis* pathogenesis. *Sci Rep*. 2022;12:19098.
91. Westphal KR, Nielsen KAH, Wollenberg RD, Møllehøj MB, Bachleitner S, Studt L, et al. Fusaoctaxin A, an example of a two-step mechanism for non-ribosomal peptide assembly and maturation in fungi. *Toxins*. 2019;11:277.
92. Yang Y, Yu L, Qiu X, Xiong D, Tian C. A putative terpene cyclase gene (*CcPtc1*) is required for fungal development and virulence in *Cytospora chrysosperma*. *Front Microbiol*. 2023;14:1084828.
93. Ferrara M, Gallo A, Perrone G, Magistà D, Baker SE. Comparative genomic analysis of ochratoxin A biosynthetic cluster in producing fungi: New evidence of a cyclase gene involvement. *Front Microbiol*. 2020;11:581309.
94. Castegnaro M, Canadas D, Vrabcheva T, Petkova-Bocharova T, Chernozemsky IN, Pfohl-Leszkowicz A. Balkan endemic nephropathy: role of ochratoxins a through biomarkers. *Mol Nutr Food Res*. 2006;50:519–29.
95. Malir F, Ostry V, Pfohl-Leszkowicz A, Malir J, Toman J. Ochratoxin A. 50 years of Research. *Toxins*. 2016;8:191.
96. Nishimura S, Tatano S, Miyamoto Y, Ohtani K, Fukumoto T, Gomi K, et al. A zinc-binding citrus protein metallothionein can act as a plant defense factor by controlling host-selective ACR-toxin production. *Plant Mol Biol*. 2013;81:1–11.
97. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing single polymerase molecules. *Science*. 2009;323(5910):133–8.
98. Kim KE, Peluso P, Babayan P, Yeadon PJ, Yu C, Fisher WW, et al. Long-read, whole-genome shotgun sequence data for five model organisms. *Sci Data*. 2014;1:140045.
99. Utturkar SM, Klingeman DM, Land ML, Schadt CW, Doktycz MJ, Pelletier DA, et al. Evaluation and validation of *de novo* and hybrid assembly techniques to derive high-quality genome sequences. *Bioinformatics*. 2014;30:2709–16.
100. Badouin H, Hood ME, Gouzy J, Aguilera G, Siguenza S, Perlin MH, et al. Chaos of rearrangements in the mating-type chromosomes of the Anther-Smut Fungus *Microbotryum Lychnidis-Dioicae*. *Genetics*. 2015;200:1275–84.
101. Sit CS, Ruzzini AC, Van Arnem EB, Ramadhar TR, Currie CR, Clardy J. Variable genetic architectures produce virtually identical molecules in bacterial symbionts of fungus-growing ants. *Proc Natl Acad Sci USA*. 2015;112:13150–4.
102. Tsuji M, Kudoh S, Hoshino T. Draft genome sequence of cryophilic basidiomycetous yeast *Mrakia Blolopsis* SK-4, isolated from an algal mat of Naga-Ike lake in the Skarvsnes ice-free area, East Antarctica. *Genome Announc*. 2015;3:e01454–01414.
103. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics*. 2008;24:637–44.
104. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 2008;18:1979–90.
105. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021;49:9077–96.

106. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35:3100–8.
107. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 2015;43:D130–7.
108. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 1999;27:573–80.
109. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nat Genet.* 2000;25:25–9.
110. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterPro-Scan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–40.
111. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44:D457–62.
112. Sperschneider DJ, Dodds DP. EffectorP 3.0: prediction of apoplasmic and cytoplasmic effectors in fungi and oomycetes. *Mol Plant Microbe Interact.* 2022;35(2):146–56.
113. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-base in 2022: a multi-species phenotype database for Pathogen–host interactions. *Nucleic Acids Res.* 2022;50:D837–47.
114. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative Toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13:1194–202.
115. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20:238.
116. Katoh K, Standley DM. MAFFT multiple sequence alignment Software Version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
117. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 2010;5:e9490.

## Publisher's Note

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