BMC Genomics

RESEARCH Open Access

Complete genome sequence analysis of *Pestalotiopsis microspora*, a fungal pathogen causing kiwifruit postharvest rots

Lei Deng[1†], Xufang Qiu[1†], Qiya Su[2], Hui Pan[1], Zupeng Wang[1], Guoliang Qian[3], Pu Liu[4], Dejiang Liu[2], Xiujun Zhang[1], Caihong Zhong[1*] and Li Li[1*]

Abstract

Background The postharvest rot of kiwifruit is one of the most devastating diseases affecting kiwifruit quality worldwide. However, the genomic basis and pathogenicity mechanisms of kiwifruit rot pathogens are lacking. Here we report the first whole genome sequence of *Pestalotiopsis microspora*, one of the main pathogens causing postharvest kiwifruit rot in China. The genome of strain KFRD-2 was sequenced, de novo assembled, and analyzed.

Results The genome of KFRD-2 was estimated to be approximately 50.31 Mb in size, with an overall GC content of 50.25%. Among 14,711 predicted genes, 14,423 (98.04%) exhibited significant matches to genes in the NCBI nr database. A phylogenetic analysis of 26 known pathogenic fungi, including *P. microspora* KFRD-2, based on conserved orthologous genes, revealed that KFRD-2's closest evolutionary relationships were to *Neopestalotiopsis* spp. Among KFRD-2's coding genes, 870 putative CAZy genes spanned six classes of CAZys, which play roles in degrading plant cell walls. Out of the 25 other plant pathogenic fungi, *P. microspora* possessed a greater number of CAZy genes than 22 and was especially enriched in GH and AA genes. A total of 845 transcription factors and 86 secondary metabolism gene clusters were predicted, representing various types. Furthermore, 28 effectors and 109 virulence-enhanced factors were identified using the PHI (pathogen host-interacting) database.

Conclusion This complete genome sequence analysis of the kiwifruit postharvest rot pathogen *P. microspora* enriches our understanding its disease pathogenesis and virulence. This study establishes a theoretical foundation for future investigations into the pathogenic mechanisms of *P. microspora* and the development of enhanced strategies for the efficient management of kiwifruit postharvest rots.

Keywords Kiwifruit, Postharvest rot, Pathogenic fungi, *Pestalotiopsis microspora*, Genome sequence analysis, CAZy genes, Transcription factors, Effectors, Virulence-enhanced factors

[†]Lei Deng and Xufang Qiu contributed equally to this work.

*Correspondence:
Caihong Zhong
zhongch@wbgcas.cn
Li Li
lili@wbgcas.cn
[1]CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, CAS Engineering Laboratory for Kiwifruit Industrial Technology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China
[2]College of Life Sciences, Jiamusi University, Jiamusi 154007, China
[3]College of Plant Protection, Nanjing Agricultural University, Nanjing 210095, China
[4]College of Horticulture, Anhui Agricultural University, Hefei 230036, China

Background

Kiwifruit is increasing in popularity worldwide owing to its high nutrient content and delicate flavor [1]. However, with the rapid expansion of cultivation areas, postharvest fungal diseases are becoming more severe, resulting in huge economic losses during the storage, transportation, marketing, and shelf-life periods [2]. Common manifestations of kiwifruit soft rot include blossom-end rot (BER; at the stylar end of the fruit), body rot (BR; on the main body of the fruit), and stem-end rot (SER; at the picking scar) [2]. *Pestalotiopsis microspora*, a fungus belonging to the family "Pestalotiopsidaceae", was initially identified as one of the main pathogenic fungi causing kiwifruit soft rot in Hubei Province, China [3]. In our previous work, *P. microspora* KFRD-2 was isolated from rotting fruits and identified by examining its morphological characteristics and ITS (internal transcribed spacer) region and beta-tubulin (BT) gene sequences together with pathogenicity testing [3, 4].

The taxonomy of *P. microspora* in the MycoBank database is as follows: Fungi, Ascomycota, Sordariomycetes, Amphisphaeriales, Pestalotiopsidaceae, *Pestalotiopsis*. *Pestalotiopsis* species are widely distributed in tropical and subtropical areas, occurring both on living plants (as pathogens or endophytes) and dead plant materials (as saprobes) [5, 6]. Increasing evidence indicates that many species of the genus *Pestalotiopsis* are common plant pathogens, causing leaf blights and post-harvest diseases like fruit rots. For instance, *P. microspora* and *P. chamaeropis* cause leaf blight diseases in *Machilus nanmu* or *Camellia sinensis*, respectively [7, 8]. Additionally, many *Pestalotiopsis* species have been identified as fruit pathogens, causing significant diseases in a diverse range of fruits and fruit plants, including blueberry dieback [9], guava scab [10], leaf blight of banana [11], fruit rot on rambutan [12], and others. On the other hand, some species have been identified as important endophytic fungi, such as the taxol-producing fungus *P. microspora* NK17 and *P. fici*, which produces 70 unique, promising natural bioactive secondary metabolites [13].

Both the lifestyle and pathogenic mechanisms of *Pestalotiopsis* fungi have not been comprehensively studied. In recent years, fungal genome sequencing has been widely applied in research on plant disease management [14], with next-generation sequencing technology facilitating genomics-based approaches. Although the whole genome of *P. fici* (https://www.ncbi.nlm.nih.gov/genome) has been reported [15], research regarding the genomic features of fungi of the Sporocadaceae family is still lacking. To facilitate the investigation of the gene regulatory network of *P. microspora*, the whole genome of strain KFRD-2 was sequenced, *de novo* assembled, and annotated in this study. Gene families encoding carbohydrate-active enzymes, especially glycoside hydrolases, are shown to have undergone expansion, and a large set of genes involved in pathogen-host interactions and secondary metabolism are identified. This genomic information may provide insights into the pathogenic mechanisms of disease development.

Results

Infection symptoms and phenotypic characteristics of KFRD-2

After in vitro incubation with KFRD-2 for 6–10 days, the typical infection symptoms on fruits were observed: slight depressions appeared on the fruit peel, then the flesh became milky white or watery, and finally, the fruits decayed severely and became sour smelling (Fig. 1a). These tests also verified strain KFRD-2 as the causative agent through Koch's postulates. After growth in potato dextrose agar (PDA) for 5 days, the mycelia of KFRD-2 appeared white and dense, with concentric annular rings (Fig. 1b). The hyphae were characterized by their septa (Fig. 1c). Conidia were slightly curved, fusiform to clavate, and five-celled, with constricted septa. They were 17.2–21.1×4.8–7.7 μm, with hyaline apical and basal cells. The apical cells had two or three 9.8–21.3 μm-long appendages. The three median cells were brown to olive green (Fig. 1d). Based on the morphological characteristics, strain KFRD-2 was preliminarily identified as *P. microspora*.

Molecular identification of KFRD-2

In our previous research [3, 4], KFRD-2 had been identified as *P. microspora* based on its ITS and β-tub sequences. However, due to the complexity of the phylogenetic relationships of *pestalotiopsis*-like fungi and because the NCBI database has been updated with abundant new data, the BLAST results for the KFRD-2 ITS accession were now different and more complicated. The closest BLAST match for the KFRD-2 β-tub sequence (Accession No. KU377338) was another *Pestalotiopsis microspora* sequence (Accession No. JN314419), with 100% identity, the same result reported in Li et al., 2017. However, the BLAST results for the KFRD-2 ITS sequence (Accession No. KR703275) were confusing, some matches with 100% nucleotide identity were *Pestalotiopsis* sp. (Accession No. EU273522), while some were *Neopestalotiopsis* sp.

In this study, we added the TEF1-α sequence of KFRD-2 (Accession No. PQ114718), and constructed phylogenetic trees to further clarify the correct taxonomy of strain KFRD-2. In single sequence analyses, β-tub seemed to perform much better than ITS and TEF1-α for species identification, strain KFRD-2 clustered with 10 *P. microspora* strains in the β-tub-based phylogenetic tree (Fig. 2a). In the phylogenetic tree constructed using concatenated ITS and β-tub sequences, strain

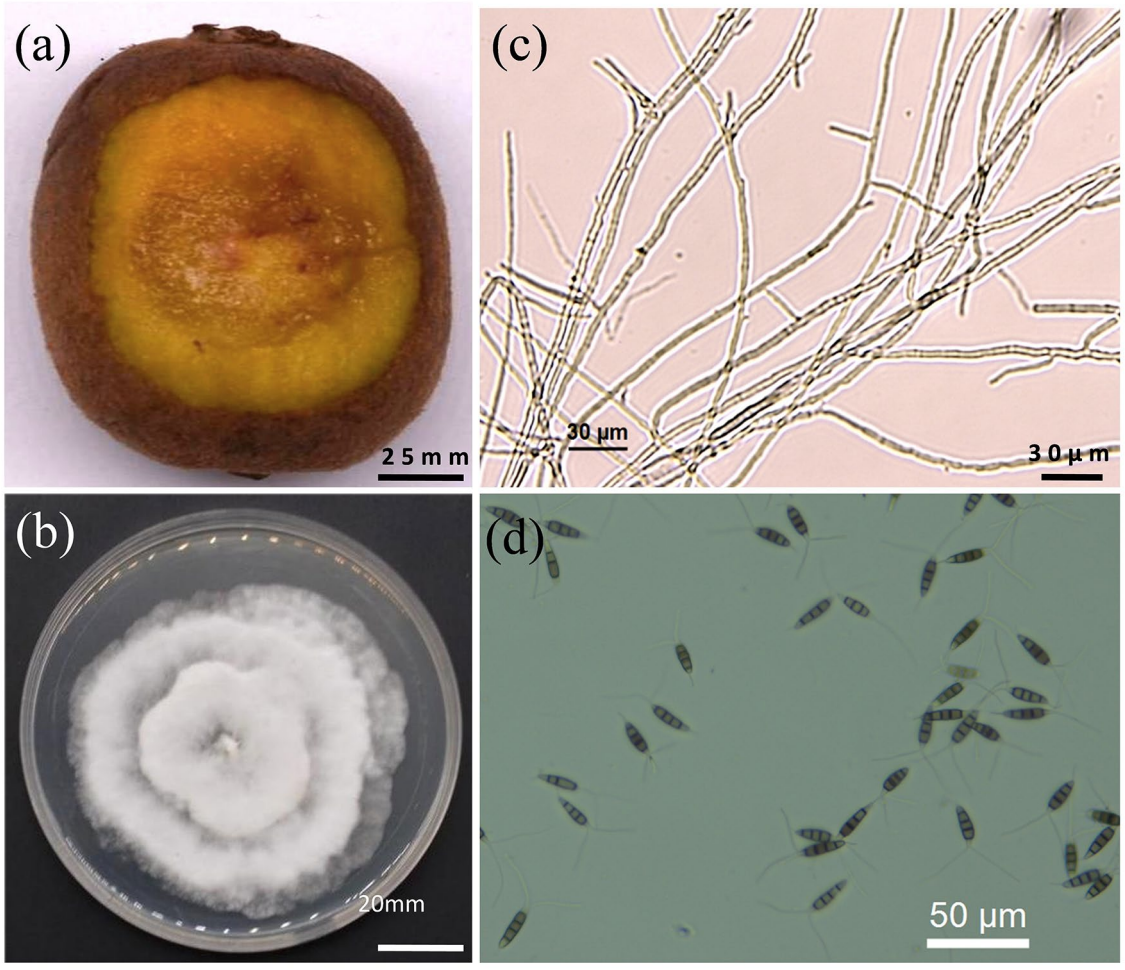

Fig. 1 Infection and phenotypic characteristics of *P. microspora* KFRD-2: (**a**) pathogenicity on the kiwifruit 'Donghong' cultivar 6 days after inoculation, (**b**) colony morphologies on PDA after 5 days of incubation, and (**c**) mycelia (scale bar = 30 μm) and (**d**) conidia (scale bar = 50 μm) under an optical microscope on the 5th day of incubation

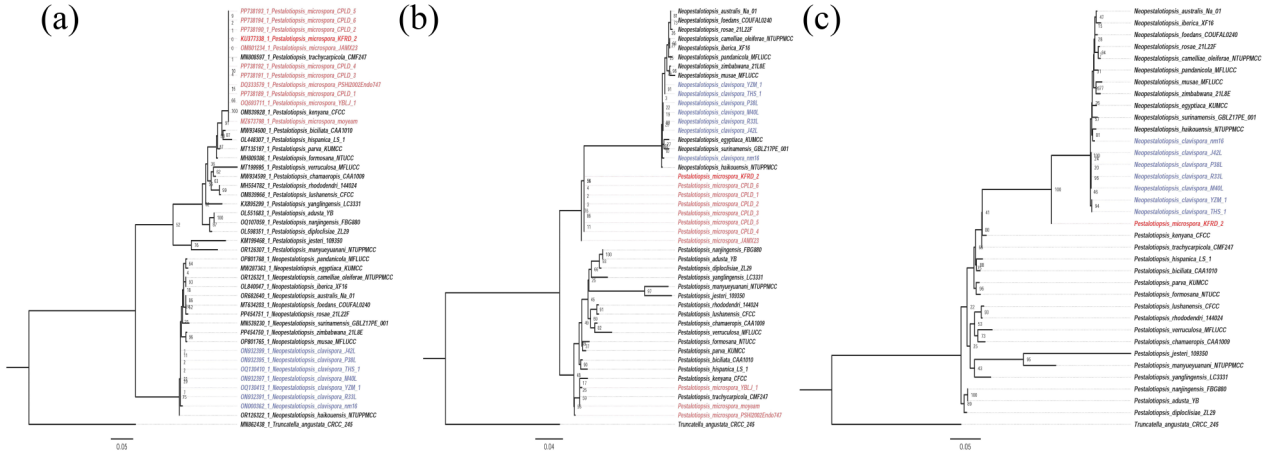

Fig. 2 Phylogenetic trees constructed by the β-tub (**a**), ITS + β-tub (**b**), and ITS + β-tub + TEF1-α (**c**) sequences, which confirm strain KFRD-2's identification as *Pestalotiopsis microspora*

KFRD-2 clustered with 7 of 10 *P. microspora* strains (Fig. 2b). Using the concatenated ITS, β-tub, and TEF1-α sequences, strain KFRD-2 was placed at an intermediate state between two genera, *Pestalotiopsis* and *Neobestalotiopsis*, in the phylogenetic tree, which reflects the continuity of species evolution. Overall, strain KFRD-2 clustered primarily with *P. microspora* strains, confirming that KFRD-2's identification as *P. microspora* is relatively correct.

General genome features

The genome of *P. microspora* was assessed using Illumina paired-end sequences data through K-mer spectrum analysis. The final de novo genome of *P. microspora* strain KFRD-2 had initial size of 50.31 Mb and an overall GC content of 50.25% (Table 1). The sequencing data statistics from both the Illumina and Pacbio platforms are listed in Table S1.

The assembled genome comprised seven contigs, with a contig N50 of 7.17 Mb. The longest contig sequence was 9.94 Mb, and the shortest was 5.83 Mb (Fig. 3; Table 1). Telomere sequences consisting of TTAGGG repeats with lengths ranging from 78 to 144 bp were identified at the head or tail end of all seven contig sequences.

Repeat elements and non-coding RNAs

The repeat sequences identified in the whole genome are listed in Table 2. The total length of the repeat sequences was estimated to be 0.73 Mb. Among the long terminal repeat (LTR) elements, Gypsy accounted for 0.16% of the assembled genome, and Copia for approximately 0.07% (Table 2). The DNA identified transposons were TcMar-Fot1 (0.13%) and MULE-MuDR (0.002%) (Table 2). The total length of 9,635 tandem repeat sequences was 424,803 bp, covering 0.84% of the genomic length (Table 2).

The analysis results for non-coding RNAs in the *P. microspora* genome are presented in Table 3. A total of 229 tRNAs and 17 rRNAs were predicted (Table 3). All tRNAs corresponded to 1 of the 20 common amino acid codons, and 159 contained introns. Additionally, three pseudogenes were identified (Table 3).

Table 1 General features of the *P. microspora* genome

Features	*P. microspora*
Genome size (bp)	50,315,141
Contig number (n)	7
Contig N50 (bp)	7,172,926
Contig N90 (bp)	5,829,117
Longest length (bp)	9,936,797
Shortest length (bp)	5,829,117
Rate of GC	50.25%

Gene prediction and annotation

A total of 14,711 protein-encoding genes were predicted. Genomic structures were analyzed, revealing a median gene length of 1,685 bp, median intergenic length of 910 bp, median complementary DNA (cDNA) length of 1,531 bp, average exon number of 2.88, and a median equal single intron length of 64 bp (Table 4). The BUSCO assessment indicated that the genome completeness, in terms of the conserved single-copy proteins of *P. microspora*, was estimated to be 99.4%, with only 18 of 3817 single-copy entries missing (Fig. 4, Table S2). Among the 14,711 identified genes, 14,439 had functional annotation results in public databases, including the nr, Swiss-Prot, Clusters of Orthologous Genes (COG), InterPro, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO) databases (Table S3). A total of 14,423 genes (98.04%) showed significant matches to genes in the NCBI nr database. Of these, nr species distribution statistics revealed that 85.56% belonged to *P. fici*, demonstrating that the sequences we produced for strain KFRD-2 were of very high quality (Table S4).

Through NCBI eukaryotic orthologous group (KOG) mapping, 4,199 proteins were assigned to KOG categories, accounting for 28.54% of the total number of genes (Fig. 5). The category "General functional prediction only" had the highest number of genes, followed by "Posttranslational modification, Protein turnover, chaperones," "Energy production and conversion," "Signal transduction mechanisms," and "Lipid transport and metabolism" (Fig. 5).

In the GO analysis, 9,399 predicted proteins, accounting for 63.89% of the entire genome, were identified. They were divided into three major subclasses: "Biological Process" (21 branches), "Cellular Component" (16 branches), and "Molecular Function" (16 branches) (Fig. 6). In the Biological Process category, the greatest numbers of genes were enriched in GO terms such as "RNA metabolic process," "cellular component organization," and "protein metabolism processes,". For cellular components, the greatest numbers of genes were enriched in "nucleus" and "protein-containing complex." For molecular biological functions, the greatest numbers of genes were enriched in functions such as "catalysis," "metal ion binding," and "small molecule binding" (Fig. 6).

Comparisons with other fungal genomes

In order to screen *P. microspora* for the presence of key virulence genes of plant pathogens, the proteome of *P. microspora* was compared with 25 sequenced pathogenic fungi known to cause major plant diseases using the NCBI and PHI databases. An OrthoMCL analysis constructed a total of 20,320 ortholog groups (OGs), and among them, 10,987 OGs contained 14,193 *P. microspora* proteins. About 3,387 (23.02%) of the predicted proteins

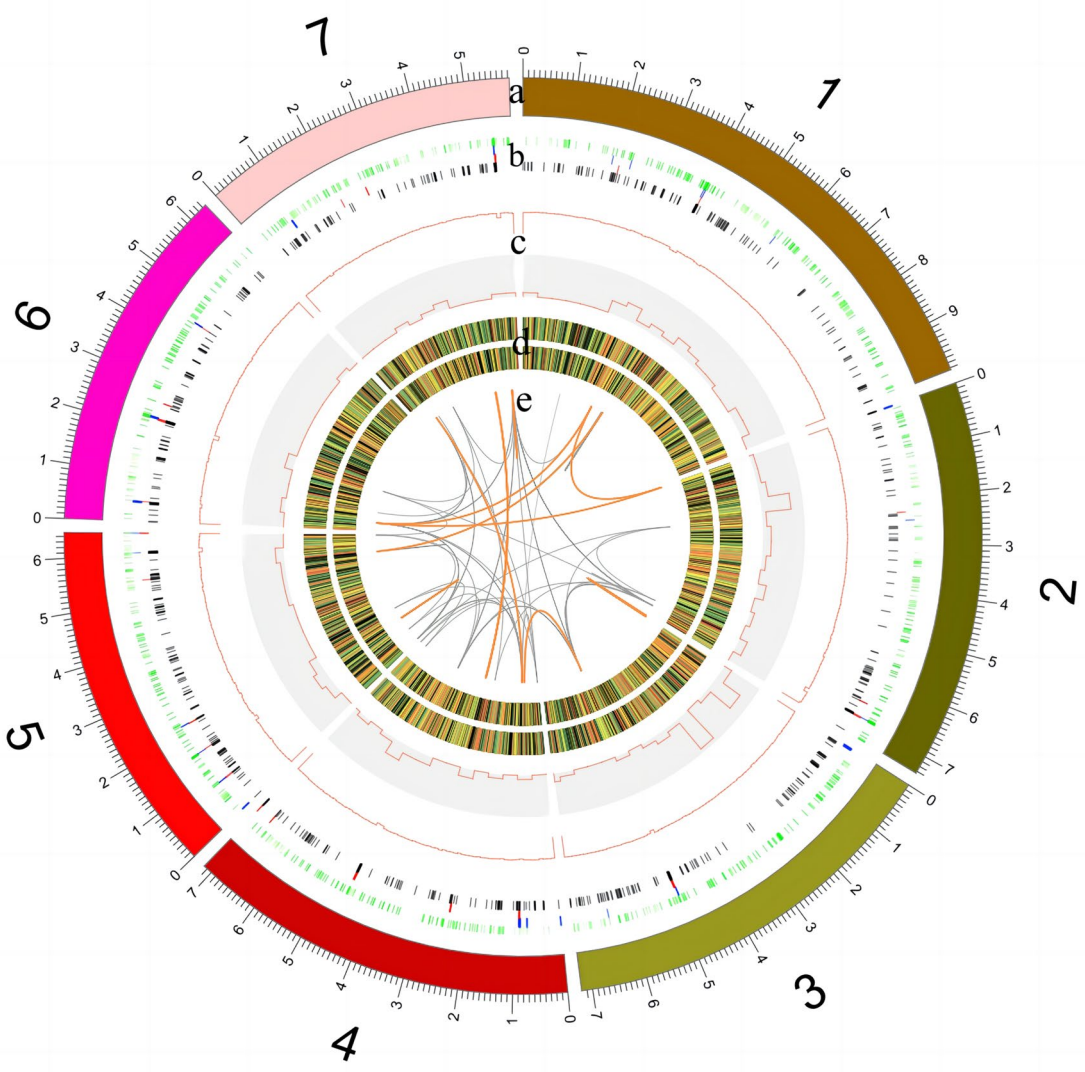

Fig. 3 Genome landscape of *P. microspora*. From the outer to inner circles: (**a**) the seven contigs of the *P. microspora* genome; (**b**) the distribution of repetitive sequences; (**c**) the GC content (outer ring) and SNP density (inner ring); (**d**) gene expression; and in (**e**), sequences sharing more than 90% identity are connected by lines, with orange lines representing sequences with lengths ≥ 10 kb and grey lines representing sequences with lengths ≥ 5 kb

Table 2 Summary of repeat sequences in the *P. microspora* genome

Repeat type		Number	Length (bp)	Percentage (%)
LTR elements				
	Gypsy	51	81,799	0.16
	Copia	23	35,829	0.07
DNA transposons				
	TcMar-Fot1	61	67,123	0.13
	MULE-MuDR	1	754	0.002
Tandem repeat		9,635	424,803	0.84

Table 3 Summary of non-coding RNAs in the *P. microspora* genome

Types	Number
rRNA	17
tRNA	229
pseudogenes	3

in *P. microspora* had orthologs in all the other species, whereas 518 (3.52%) were unique to *P. microspora*, approximately 1.74% of which had at least one paralog. A phylogenetic tree was constructed using all the single-copy orthologous genes conserved in the 25 fungi fungal pathogens, and *P. microspora* clustered with other *Xylariales* species, with the closest relative being *Neopestalotiopsis* sp. (Fig. 7).

Table 4 Genome annotation statistics for the *P. microspora* genome

Item	Median	Length (bp)
Gene_length	1685	1931.4235
Intergenic_length	910	1613.9765
cDNA_length	1531	1764.8743
CDS_length	1272	1478.9690
Intron_length	114	164.2666
Single_exon_length	372	613.0015
Single_CDS_length	294	520.2312
Single_intron_length	64	87.4191

As the whole genome of *P. fici* is publicly available, the protein sequences of *P. microspora* and *P. fici* were compared, revealing 98 blocks of covariance. At least 5 genes in each block were co-linear on the genome. These co-linear blocks contained a total of 11,119 genes from either *P. microspora* or *P. fici*. Among the co-linear blocks, 79 had more than 10 genes, with the largest co-linear block containing 1204 genes (Fig. 8). The genetic diversity of the genus *Pestalotiopsis* could be further investigated to provide an important basis for understanding the pathogenic mechanism of *P. microspora* in kiwifruit soft rot.

Transcription factors

A total of 43 types of transcription factors were identified in the *P. microspora* genome. Among the 845 individual transcription factor genes identified in the strain, the highest number of genes were identified as zf-clus (320), followed by fungal-trans (237) and zf-C2H2 (101) (Table S5). These three types of transcription factors accounted for 77.87% of the total number of transcription factors (Fig. 9).

Carbohydrate-active enzymes

Microbe-derived carbohydrate-active enzymes (CAZys) are responsible for the degradation of plant cell wall carbohydrates. For fungi, CAZys can be exploited to destroy the cell wall structural integrity of the host and acquire nutrients [16]. Based on their sequences and structural similarities within the functional domains, CAZys can be classified into six main classes: glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide

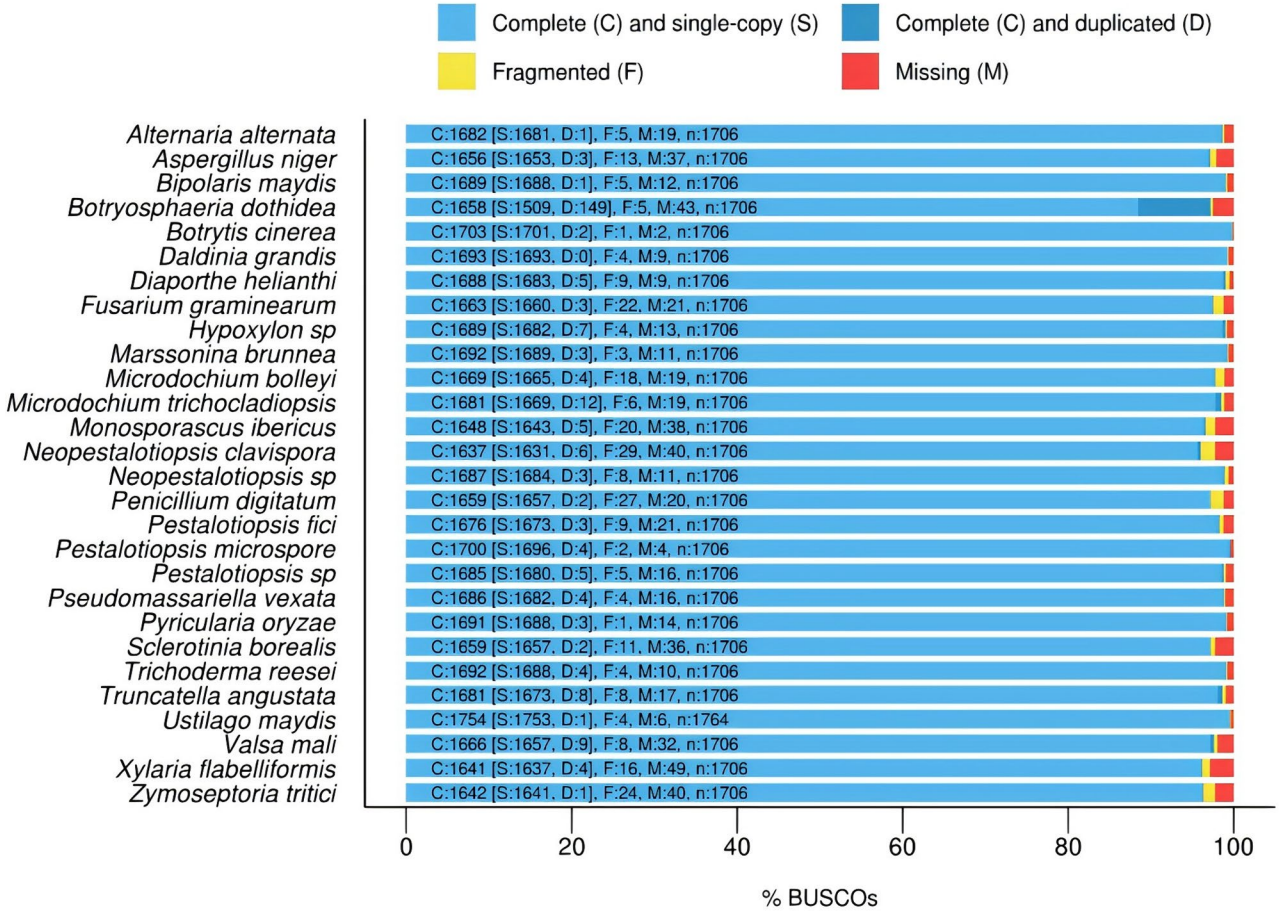

Fig. 4 BUSCO assessments of the *P. microspora* genome and other pathogenic fungi genomes. Genome completeness was estimated to be 99.4%

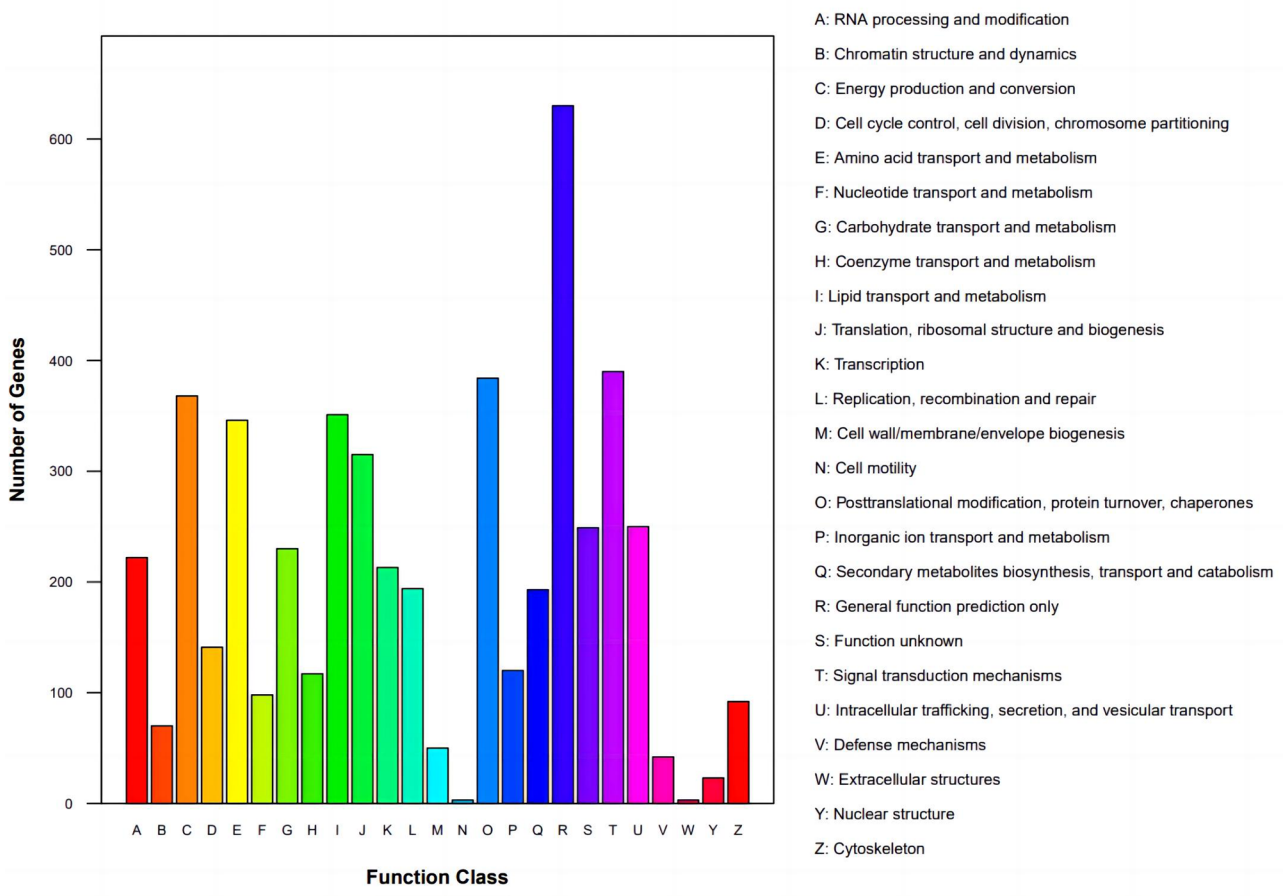

Fig. 5 KOG function classification clusters for the proteins identified in the *P. microspora* genome. Letters along the abscissa represent the functional classification from the KOG analysis, and the ordinate displays the number of genes

lyases (PLs), glycosyltransferases (GTs), auxiliary activity enzymes (AAs), and carbohydrate-binding modules (CBMs) [17]. A total of 870 putative CAZy genes were identified in *P. microspora* by mapping genes against the CAZy database. Among these, 395 genes encoded GHs, 105 encoded GTs, 79 encoded CBMs, 82 encoded CEs, 32 encoded PLs, and 236 encoded AAs (Table S6). Comparing the numbers of CAZy genes in *P. microspora* with those of 25 other pathogenic fungi, *P. microspora* was found to have a greater number of CAZy genes than 22 of the other species, especially in the number of GH and AA genes (Fig. 10). This list of *P. microspora*'s CAZys lays the foundation for dissecting the mechanisms of its fungal pathogenicity.

Virulence genes

A total of 5,465 pathogen host-interacting (PHI) genes, including 28 effectors, 102 virulence-enhancing genes, and 109 lethal factors, were found in *P. microspora* by comparing its genes to the PHI-base database (Fig. 11, Table S7). Annotations of the 28 predicted effector genes are listed in Table 5.

Secondary metabolite-associated gene clusters

Gene clusters involved in the secondary metabolism of *P. microspora* are shown in Table S8. There were 86 gene clusters identified in the *P. microspora* genome, including 31 T1 polyketide synthase (PKS), 15 Terpene, 10 nonribosomal peptide synthase (NRPS), 4 Indole, 3 T1PKS-NRPS, 1 Terpene-NRPS, 1 T3PKS, 1 Indole-T1PKS, 1 lantipeptide, and 19 other gene clusters (Fig. 11). Among these, 47 PKS and NRPS gene clusters were revealed, accounting for 54.65% of the total number of predicted gene clusters (Fig. 12).

Discussion

In recent years, based on the conidial color of their median cells and multi-locus molecular phylogenies, the *pestalotiopsis*-like fungi were classified into three genera: *Pestalotiopsis*, *Pseudopestalotiopsis*, and *Neopestalotiopsis.* Several species originally classified under *Pestalotiopsis* were transferred to *Neopestalotiopsis* or *Pseudopestalotiopsis* [18, 19]. According to the taxonomy provided by NCBI, *Pestalotiopsis microspora* is still classified as "Ascomycota, Sordariomycetes, Xylariales, Sporocadaceae, *Pestalotiopsis*". In our manuscript, strain

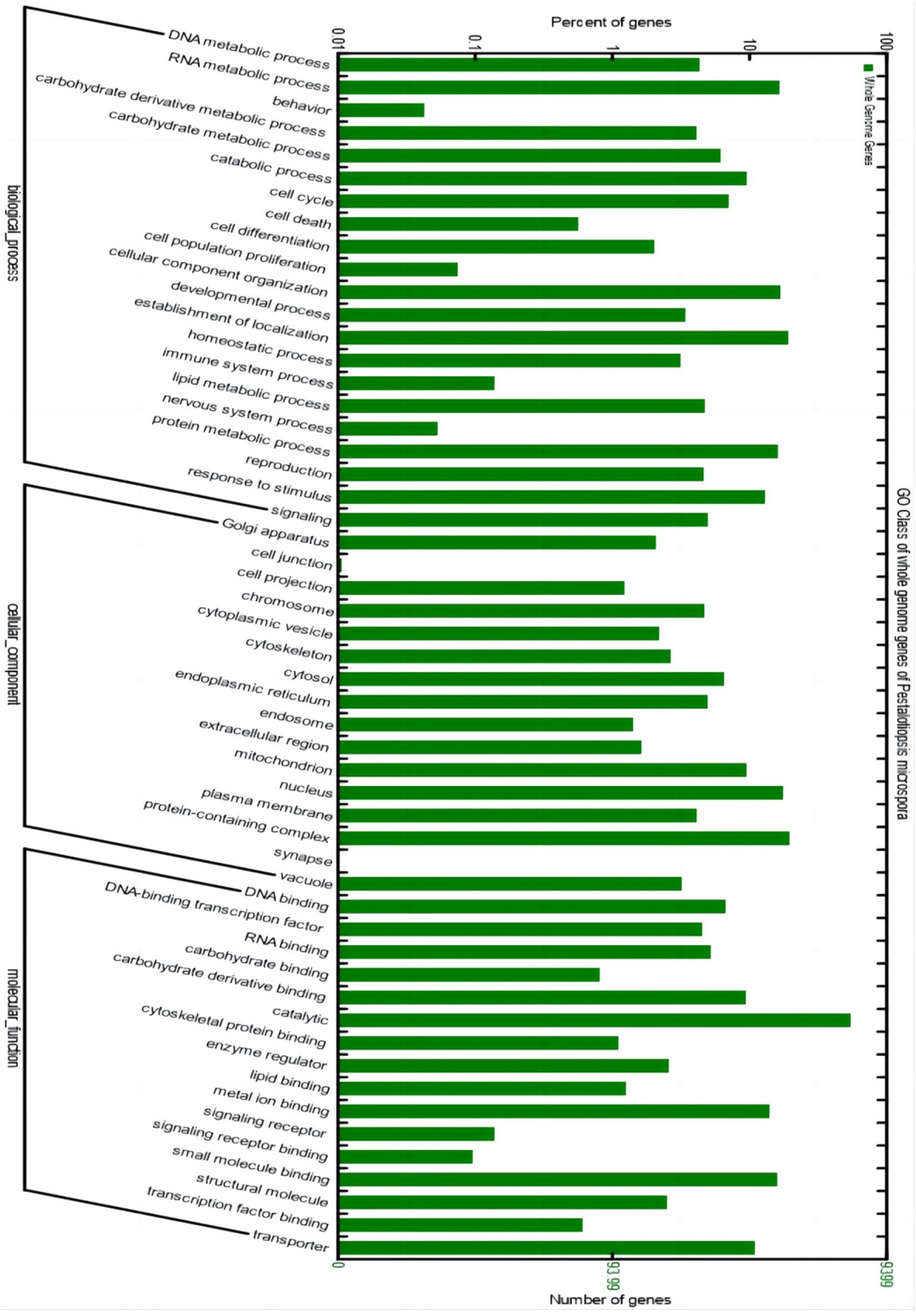

Fig. 6 Gene Ontology (GO) functional annotations of *P. microspora* genes. A total of 9,399 predicted proteins, accounting for 63.89% of the entire genome, were identified and categorized into three major subclasses: "Biological Process," "Cellular Component," and "Molecular Function"

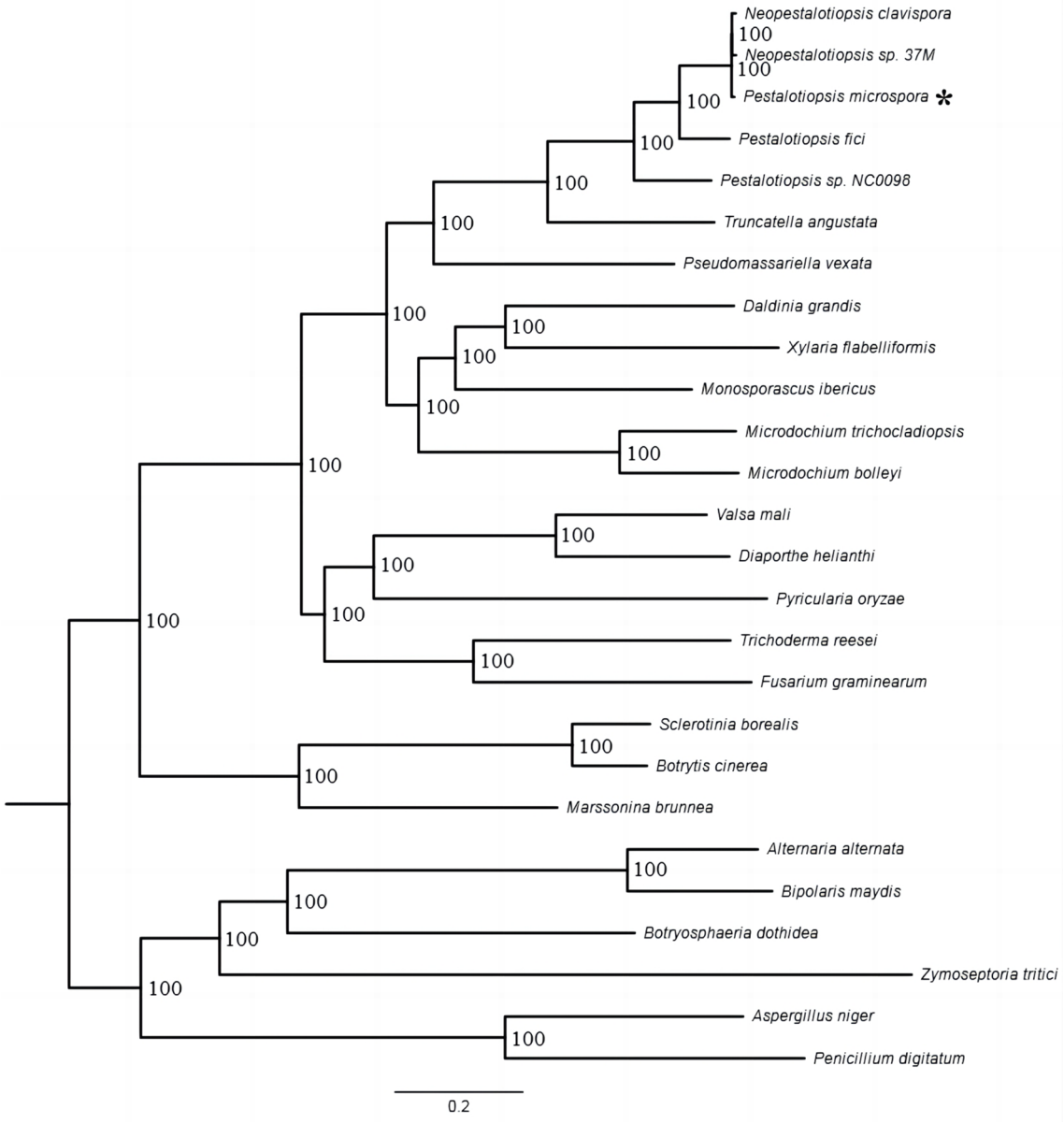

Fig. 7 Phylogenetic tree of *P. microspor*a and 25 other fungal pathogen species based on the single-copy orthologous genes conserved in these 25 fungi. The topology of the phylogenetic tree was constructed using the maximum likelihood method (LG + I + G + F model), and all bootstrap values, based on 1000 bootstrap resamples, are 100%. The time scale is shown in MYA (million years ago)

KFRD-2 was identified as *P. microspora* according to its morphological characteristics, pathogenicity tests, and a phylogenetic analysis constructed using the concatenated ITS, β-tub, andTEF1-α sequences.

Further, the general genome features of *P. microspora* KFRD-2, a *P. microspora* strain from China, were described. The genome assembly resulted from an analysis combining the high-throughput Illumina HiSeq 4000 system and PacBio sequel sequencing platform. A de novo genome assembly was generated and characterized. The *P. microspora* genome was estimated to be approximately 50 Mb, using both kmer and read coverage analyses. This genome size is comparable to that of other reported kiwifruit fungal pathogens, such as *Diaporthe phragmitis* NJD1 (55.6 Mb) [20] and *Alternaria alternata* Y784-BC03 (32.30 Mb) [21]. The overall number of predicted genes in *P. microspora* was 14,711, while they were 12,393 and 12,835 in *D. phragmitis* NJD1 and *A. alternata* Y784-BC03, respectively. The association between the number of predicted genes and the level of fungal specialization, as well as pathogenicity against kiwifruit, remains unknown. It has been reported that *Diaporthe* species collected from the 11 main kiwifruit cultivation regions of China had the highest identification rates and

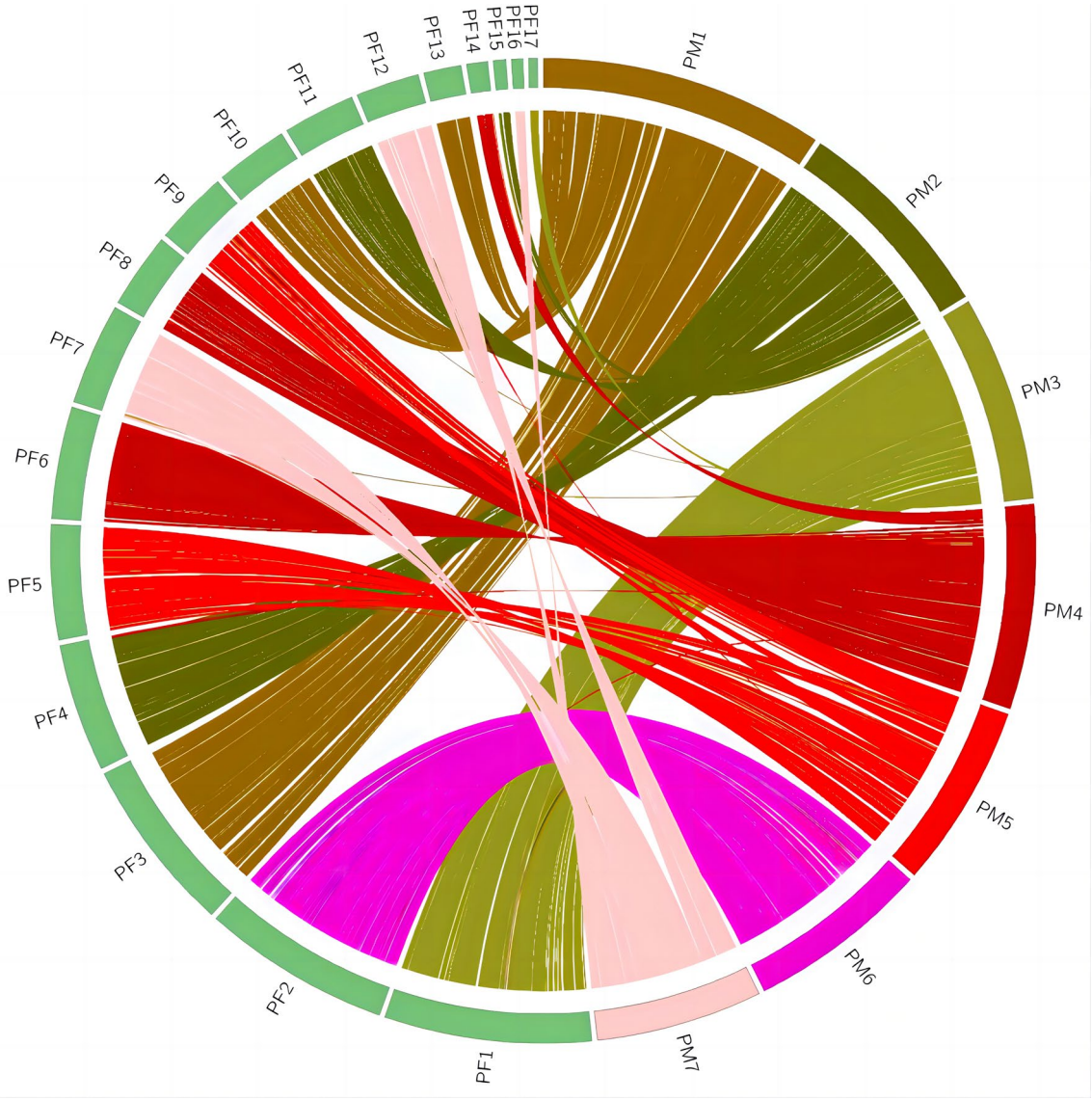

Fig. 8 Genome collinearity between *P. microspora* (right side, multicolored contigs) and *P. fici* (left side, light green contigs), showing synteny between the two genomes. Colored lines indicate homologous genes shared between syntenic blocks (containing at least 10 orthologous genes)

pathogenicity on kiwifruits during the 2014–2015 period [4, 22]. Further comparative genome analyses can be carried out to address questions about the pathogenetic mechanisms of different pathogen species.

To examine the genome architecture of *P. microspora*, repetitive elements were analyzed. Approximately 1.5% of the *P. microspora* genome consisted of repetitive elements. In filamentous fungi genomes, repetitive elements can be divided into two types: class I and class II [23]. Class I elements (retrotransposons) are transcribed from DNA to RNA and then reverse transcribed into DNA. Moreover, retrotransposons mobilize throughout the genome in a "copy-and-paste" manner [24]. In addition, long terminal repeats (LTRs) are one of the main classifications of retrotransposons [25]. The two main superfamilies of LTR retrotransposons found in fungi are Gypsy and Copia. Among the repetitive elements in *P. microspora*, 30.7% of the LTRs were identified as Copia superfamily retrotransposons and 68% as Gypsy. Class II elements (DNA transposons) are excised from the genome and integrated elsewhere. Thus, DNA transposons mobilize via a "cut-and-paste" mechanism [26]. Interestingly, only two types of DNA transposons were identified in the *P. microspora* genome: TcMar-Fot1 and MULE-MuDR. The function of these transposons in *P. microspora* is unknown.

The plant cell wall serves as a significant barrier to the invasion of plant pathogens. A considerable amount of fungal pathogens must first breach the plant cell wall before colonizing the host [27]. More than 90% of the

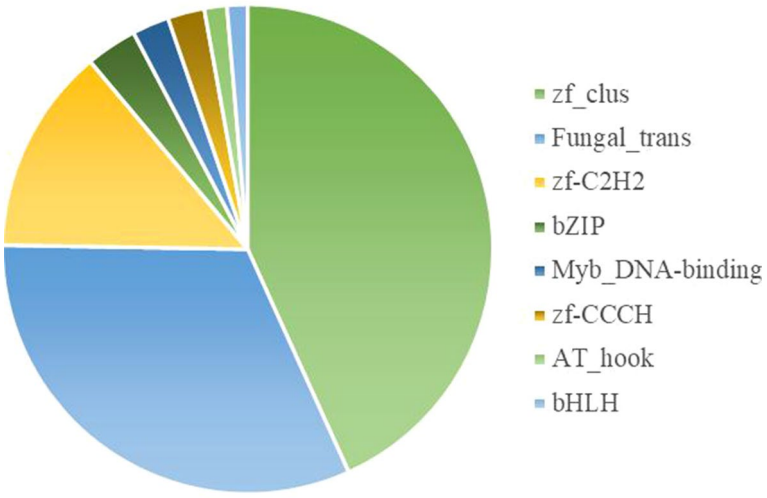

Fig. 9 Identities of transcription factors in the *P. microspora* genome. A total of 845 transcription factor genes were identified, with zf-clus, fungal-trans, and zf-C2H2 accounting for 77.87%

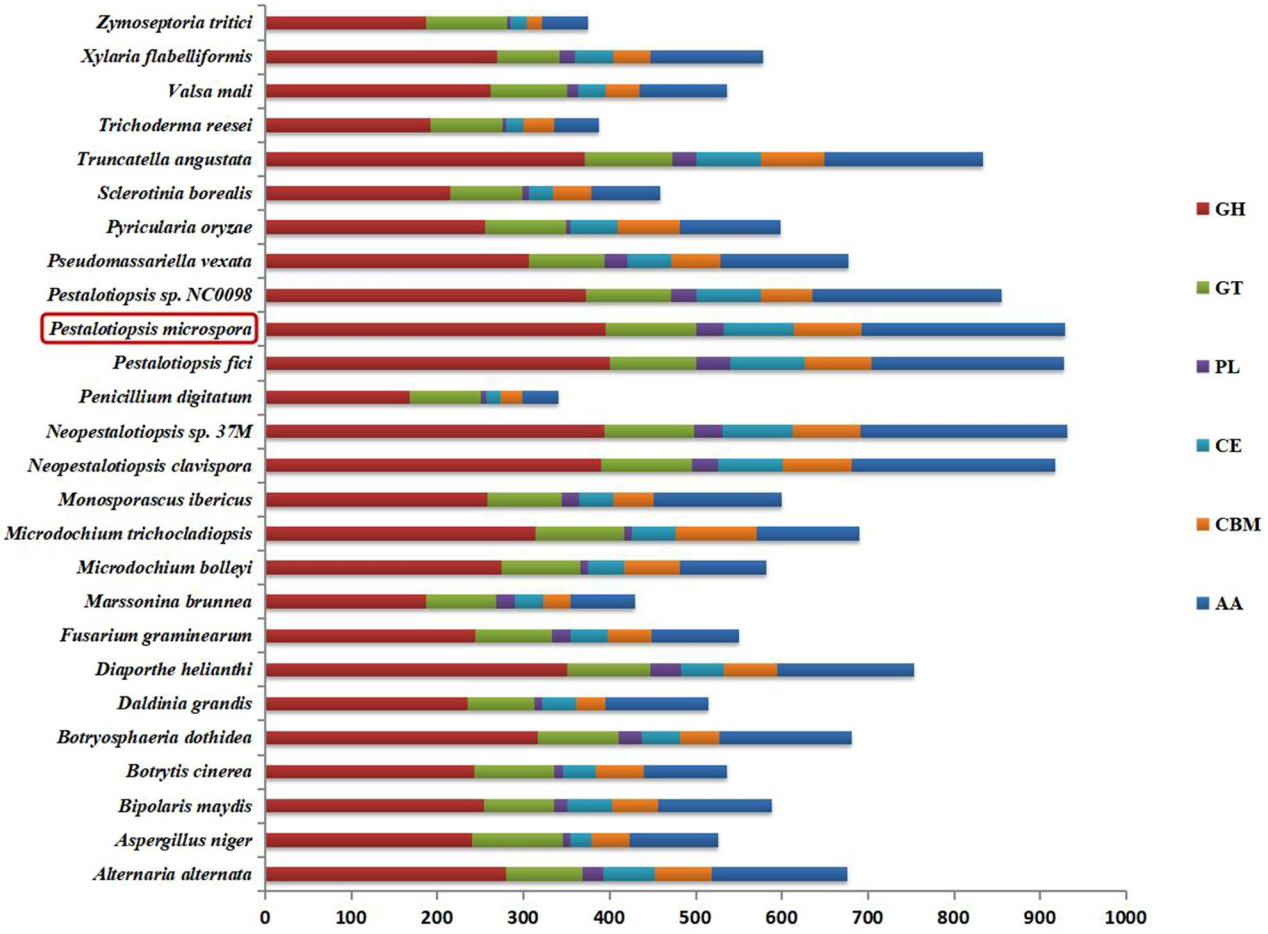

Fig. 10 The number and type of CAZys in *P. microspora* and 25 other pathogens. By mapping *P. microspora*'s genes against the CAZy database, a total of 870 putative CAZy genes were identified and classified into the six main CAZy classes (colors)

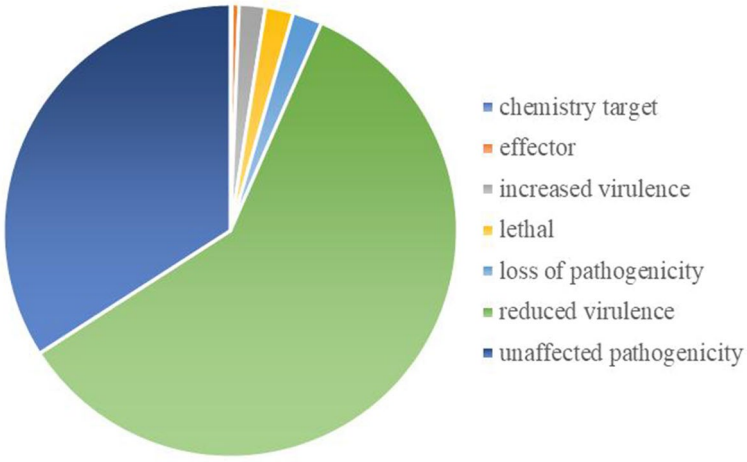

Fig. 11 Classifications of pathogen host-interacting (PHI) genes identified in the *P. microspora* genome. A total of 5,465 PHI genes were characterized by comparing them to those in the PHI-base database

plant cell wall is comprised of carbohydrates, cellulose, hemicelluloses, and pectic polysaccharides [28]. Microbe-derived CAZys have been revealed to be a large, complex, and redundant enzyme system for the degradation of plant cell walls, serving as important virulence factors for fungal pathogens [16]. To investigate the genetic basis of pathogenicity, CAZys in *P. microspora* KFRD-2 were identified and annotated. *Pestalotiopsis microspora* had among the highest number of putative CAZy genes and abundant CAZy families when compared to 25 other pathogenic fungus genomes (Table S6). The total CAZy repertoire for *P. microspora* was similar to that of other *Pestalotiopsis* species, *Neopestalotiopsis* species, and a species of the related genus *Truncatella* (Table S6). Interestingly, these fungi (*Neopestalotiopsis clavispora* and *Truncatella angustata*) are well-known pathogens of plants [29, 30].

Pectins, with various modifications, are a major component of plant cell walls. Microbial pectinolytic enzymes have been extensively studied [31, 32]. Polysaccharide lyases (PLs) specialize in pectin degradation. Furthermore, pectin can also be a substrate for several glycoside hydrolases (GHs) [33]. Compared with other fully sequenced fungi, *P. microspora* had a higher number of candidate pectinases, covering all pectinase families known from fungi, including PL1, PL3, PL4, PL9, GH28, GH78, GH88, GH95, GH105, and GH115 (Table S6). The predominant families of pectinases in the *P. microspora* genome were PL1 and GH28, with 17 and 20 encoding genes, respectively (Table S6).

The number and type of secreted GH proteins are highly variable in plant-associated fungi and oomycetes with different lifestyles [34]. Compared with necrotrophic and hemibiotrophic fungi, biotrophic pathogens and endophytic microorganisms produce relatively low numbers of GHs, minimizing host damage [35, 36]. Many necrotrophic and hemibiotrophic fungi, like *Fusarium* species, have around 300 GH-encoding genes [34]. The genome of *P. microspora*, whose lifestyle is largely unknown, encoded 395 GHs. Such a large number of GH-encoding genes indicates that *P. microspora* probably acts as a necrotroph or hemibiotroph, relying on limited nutrients within plant tissues. Enzyme substrates for the GH1, GH3, and GH5 include not only cellulose, but also hemicellulose and pectin [33], and it has been reported that most biotrophic fungi do not possess GH1 [34]. Moreover, both GH3 and GH5 are more common in hemibiotrophic and necrotrophic fungi [37]. Notably, there are 4 GH1, 42 GH3, and 22 GH5 genes in the *P. microspora* genome.

The web-accessible PHI-base database catalogs experimentally verified pathogenicity, virulence, and effector genes from fungal, oomycete, and bacterial pathogens. Based on our PHI analysis, several genes related to virulence and effector genes were predicted in the *P. microspora* genome. In recent years, studies have increasingly focused on the function of phytopathogenic fungal effectors. Phytopathogenic fungi share a strategy of secreting versatile effectors to target and modulate host immunity [38]. In this study, 28 effectors were identified in *P. microspora* at the genomic level, and they were predicted to be key factors in pathogen-host interactions. Among them, 12 were CAZys (GH12, CBM50, and AA). Interestingly, the GH12 family gene *PM01Gene04859*, encoding Xyloglucan-specific endo-beta-1,4-glucanase, had a high similarity with the *PsXEG1* gene, a well-known pathogen-associated molecular pattern (PAMP) of *Phytophthora sojae* [39]. Both *PM01Gene02998* and *PM01Gene03346* were annotated as NLPs (necrosis and ethylene-inducing-like proteins) with a putative npp1 domain. These NLP proteins can trigger cell death, phytoalexin, ethylene synthesis, and defense gene transcription activation

Table 5 Putative effectors, annotated by PHI prediction

Gene ID	InterPro annotation	CAZys
PM01Gene00152	IPR018392: LysM domain	CBM50
PM01Gene00331	IPR005103: Glycoside hydrolase, family 61	AA9
PM01Gene02511	IPR003812: Fido domain	-
PM01Gene02998	IPR008701: Necrosis inducing protein	-
PM01Gene03346	IPR008701: Necrosis inducing protein	-
PM01Gene04118	IPR018392: LysM domain	CBM50
PM01Gene04285	IPR000073: Alpha/beta hydrolase fold-1	-
PM01Gene04805	IPR018392: LysM domain	-
PM01Gene04859	IPR002594: Glycoside hydrolase family 12	GH12
PM01Gene04976	IPR000994: Peptidase M24 subfamily 1	-
PM01Gene05001	IPR001384: Peptidase M35, deuterolysin	-
PM01Gene07890	IPR019901: Ergot alkaloid biosynthesis protein	-
PM01Gene08822	IPR001002: Chitin-binding, type 1	CBM50
PM01Gene09454	-	-
PM01Gene09846	IPR006861: Hyaluronan/mRNA-binding protein	-
PM01Gene09871	IPR018392: LysM domain	CBM50
PM01Gene09967	IPR018392: LysM domain	-
PM01Gene10370	IPR018392: LysM domain	CBM50
PM01Gene10603	IPR018392: LysM domain; IPR036779	CBM50
PM01Gene10606	IPR018392: LysM domain; IPR036779	CBM50
PM01Gene11109	IPR008258: Transglycosylase SLT domain 1	-
PM01Gene11684	IPR000994: Peptidase M24 aminopeptidase	-
PM01Gene13080	IPR000073: Alpha/beta hydrolase fold-1	-
PM01Gene13322	IPR018392: LysM domain	CBM50
PM01Gene13625	-	-
PM01Gene13871	IPR000994: Peptidase M24	-
PM01Gene14220	IPR018392: LysM domain	CBM50
PM01Gene14634	IPR002594: Glycoside hydrolase family 12	GH12

in numerous dicotyledons [40]. In our study, some LysM domain-containing proteins were also identified. The LysM domain directly binds to chitin oligosaccharides, as has been reported for the chitin recognition receptor OsCEBiP. Thus, a pathogen could deliver apoplastic effector proteins (AEPs) with LysM domains into the host that impair PAMP-triggered host immunity. For example, the fungus *Cladosporium fulvum* effector Ecp6 sequesters chitin oligosaccharides and suppresses chitin-triggered host immunity, facilitating the establishment of infection [41]. These LysM domain-containing proteins may be good candidate virulence factors for further study (Table 5).

Conclusion

The genome of strain KFRD-2 represents the first reported genome sequence of the fungus *P. microspora*, a causative agent of kiwifruit soft rot [4]. The genome contained a number of unique features, including an abundance of genes encoding transcription factors, cell-wall degrading enzymes, and PHI. Additionally, *P. microspora* strain KFRD-2 demonstrates the potential to synthesize a variety of secondary metabolites and gene components. These discoveries enhance our knowledge about the genomic biology and gene regulatory network of *P. microspora*. In summary, the genomic sequence information obtained in this study establishes a theoretical foundation for future research on the interactions between *P. microspora* and kiwifruit.

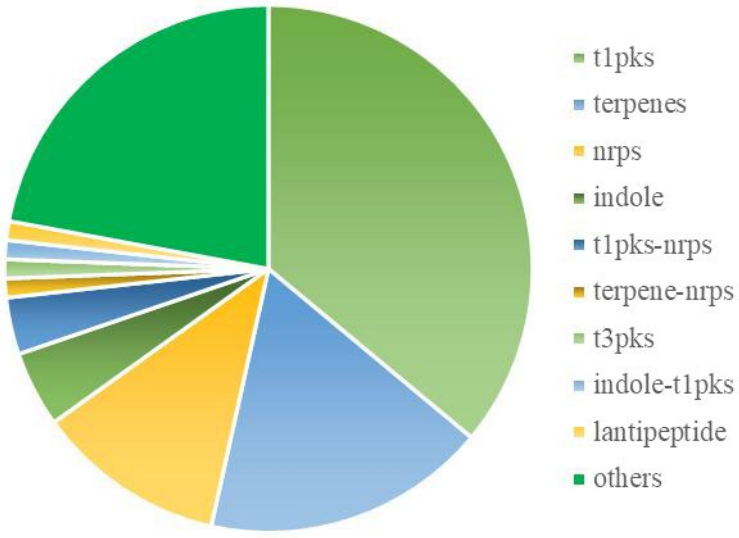

Fig. 12 Secondary metabolite-associated gene clusters identified in the *P. microspora* genome. A total of 86 gene clusters were predicted using the SMURF platform

Materials and methods

Strains and culturing conditions

The *P. microspora* strain KFRD-2 was originally isolated from diseased fruits of the *Actinidia chinensis* 'Jinyan' cultivar, which were purchased from the main kiwifruit-cultivating regions in Wuhan (Hubei Province, China). The strain was deposited at the Engineering Laboratory for Kiwifruit Industrial Technology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, PR China. For pathogenicity testing, a mycelial plug (5 mm) of actively growing *P. microspora* was rubbed onto mature 'Donghong' fruits, which were then incubated at 22/18°C (day/night) and 80% humidity for 10 days. For morphological observations, cultures were grown in PDA medium (200 g of potato extract, 20 g of dextrose, and 15 g of agar per liter water) at 25 °C for 5 days with constant light. Taxonomical identification of *P. microspora* was based on morphological characteristics and an analysis of the ITS region of the rDNA and the BT gene sequences [3, 4].

Genomic DNA extraction and high-throughput sequencing

For whole genome sequencing, *P. microspora* strain KFRD-2 was cultured on solid PDA medium. After six days of incubation at 25℃ in the dark, mycelia were collected and immediately frozen with liquid nitrogen for genomic DNA extraction using a Qiagen DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA). The whole genomic DNA sample was used to generate sequencing libraries, and all libraries were sequenced on both the Illumina Hiseq 4000 and PacBio sequel platforms by Novogene Biotech AG (Tianjing, China).

Genome survey and assembling

The adaptor sequence and low-quality bases of the paired-end Illumina raw reads were truncated using Trimmomatic v0.38 software [42]. Next, the duplicated reads derived from PCR or optical photography were removed using FastUniq version 1.1 software. The obtained clean reads were then used for a genome survey based on the k-mer algorithm and genome sequence correction.

The subread data from the PacBio Sequel sequencing platform was output in the BAM format, and the sequence files were converted into FASTA format using the "fasta" command of Samtools v1.9 software for subsequent genome assembly. Three popular *De novo* methods, including those of the Canu v2.1.1, NextDenovo v2.4.0 (https://github.com/Nextomics/NextDenovo), and FALCON v1.8.1 programs, were separately used for genome assembling [43, 44], and the quickmerge v0.3 software was used to combine the three genome assemblies. Next, the Illumina reads were aligned to the genome assembly using bowtie2 v2.4.1, with the aim of removing false positive sequences with lower coverage [45]. Then, the ARROW algorithms of the SMRTanalysis v2.3.0 and Pilon v1.23 programs were used to perform three rounds of correction on the genome sequence. Finally, the homolog_genewise *Lentinula module* within the GETA v2.7.1 pipeline (https://github.com/chenlianfu/geta) was used to identify and remove the mitochondrial genome sequences, employing *Lentinula edodes* as a reference genome. Subsequently, Perl scripts were employed to organize and rename the genome sequences.

Genome feature annotation

The transposable elements were masked using RepeatMasker v4.0.7 and RepeatModeler v1.0.9 software [46]. The coding genes were predicted with the software GETA v2.4.2, using the default parameters [21]. Unigenes were identified by querying the NCBI non-redundant nucleotide sequence (Nr) database using BLASTN v2.6.0 with a cut-off E-value of $1e^{-5}$ [47]. The gene model was acquired by comparing the protein sequences from strain KFRD-2 with those of 25 pathogenic species downloaded from NCBI. The HMMER v3.1b2 program was executed using the AUGUSTUS v3.3 software for training and gene prediction [48, 49]. The telomere regions, which consisted of TTAGGG repeats, were identified by a custom Perl script. To test the genome's integrity, all the protein sequences and nucleic acid sequences were respectively compared with the 3817 single-copy homologous genes in the sordariomyceta_odb database using the BUSCO v3.0.2b software [50]. All the predicted genomic information was used to draw a circular genome map with the CIRCOS v0.67 tool [51].

Gene function annotation

Gene functions were predicted according to the best matches of alignments against the Nr v20210824 [52] and Swiss-Prot v20210824 [53] database sequences using BLASTP searches (E-value=$1e^{-5}$) [54]. The eggnog-mapper software was used to get eggNOG v2.1.5 annotations based on the eggNOG v4.5 database and HMMER database [55]. A Perl program provided by EBI (https://www.ebi.ac.uk/) was used for InterPro annotations. Subsequently, GO enrichment analyses were carried out to obtain functional annotations. To determine whether these genes might participate in any functional pathways, all gene models were aligned (*E*-value=$1e^{-5}$) to the KO database. The functions of these putative genes were predicted and classified using the KOG database [56]. The online KEGG Automatic Annotation Server was used to assign the assembled sequences to KEGG pathways [57].

Species tree construction and collinear analysis

A comparative genomic analysis was carried out using the genomes of *P. microspora* and 25 other pathogenic fungi species. The genome data of the other 25 fungal species were downloaded from the NCBI database (Table S9). Orthologous gene relationships among species were determined using OrthMCL v2.0.9 and reciprocal BLASTp search results with identities>50% and query coverages>50%. The obtained results were analyzed using OrthoMCL with the default parameters to obtain the orthologous genes. All single-copy homologous genes for each species were extracted and aligned with MAFFT v7.221 software. Conserved block regions in each gene alignment were picked out using Gblocks v0.91b with the software's default parameters. These blocks were then concatenated into a long sequence for each species. Maximum likelihood topology searches were performed in RAxML v8.1.24 using the "PROTGAMMALGX" model, and the analysis was conducted with 1,000 bootstrap resamples. The resulting tree was visualized using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/). In addition, collinear genes between the *P. microspora* genome and adjacent species were analyzed with the Multiple Collinearity Scan Toolkit (MCScanx) [58].

CAZy family analysis

Using the CAZy database (http://www.cazy.org) as a reference, the protein sequences of all CAZy family genes were predicted using dbCAN v6.0 software [59]. The HMM domain information for the entire CAZy family was obtained using the HMM algorithm. On this basis, `.

PHI prediction and secondary metabolism gene cluster analysis

Genes associated with pathogen–host interactions were predicted via the PHI v4.8 database (http://www.phi-base.org/). To count the number of various types of genes, BLASTP was used to compare the predicted protein sequences to the whole genome's protein sequences. Secondary metabolic gene clusters were predicted using the web-based tool SMURF through the antiSMASH v5.1.2 pipeline [60, 61].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10751-y.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8
Supplementary Material 9

Acknowledgements

We thank Dr. Lianfu Chen for his excellent technical support.

Author contributions

L.D., analyzed data, executed software, and drafted the manuscript; X.Q. revised manuscript drafts and updated partial figures; Q.S., H.P., and Z.W. curated data and performed the formal analysis; G.Q., P.L., D.L., and X.Z. improved the manuscript; C.Z. administrated and supervised the project and provided partial funding support; L.L. designed and conceived the experiment, acquired funding, and revised the manuscript. All authors have reviewed and approved the manuscript.

Funding
This study was financially supported by the National Natural Science Foundation of China (32272506; 31901980; 32402471), National Key Research and Development Program of Hubei province (2022BBA0076), Funding of Hubei Province's Industrial Technology System (2023HBSTX4-08), and Hubei Hongshan Laboratory (2021hszd017).

Data availability
The genome raw sequencing data and the assembly reported in this paper is associated with NCBI BioProject accession number PRJNA1059488, and BioSample: SAMN39206026 within GenBank. The accession number of assembled genome have been assigned JBFTYK000000000. The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article and supplemental materials.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 6 January 2024 / Accepted: 30 August 2024
Published online: 06 September 2024

References

1. Ma T, Lan T, Ju Y, Cheng G, Que ZH, Geng T, et al. Comparison of the nutritional properties and biological activities of kiwifruit (*Actinidia*) and their different forms of products: towards making kiwifruit more nutritious and functional. Food Funct. 2019;10(3):1317–29.
2. Manning M, Burdon J, DeSilva N, Meier X, Pidakala P, Punter M, et al. Maturity and postharvest temperature management affect rot expression in 'Hort16A' kiwifruit. Postharvest Biol Tec. 2016;113:40–7.
3. Li L, Pan H, Chen M, Zong C. First report of *Pestalotiopsis microspora* causing postharvest rot of kiwifruit in Hubei Province, China. Plant Dis. 2016;100(10):2161.
4. Li L, Pan H, Chen M, Zhang S, Zhong C. Isolation and identification of pathogenic fungi causing postharvest fruit rot of kiwifruit (*Actinidia chinensis*) in China. J Phytopathol. 2017;165(11–12): 782 – 90.
5. Maharachchikumbura S, Guo L, Cai L, Chukeatirote E, Wu W, Sun X, et al. A multi-locus backbone tree for *Pestalotiopsis*, with a polyphasic characterization of 14 new species. Fungal Divers. 2012;56(1):95–129.
6. Maharachchikumbura S, Guo L, Chukeatirote E, Bahkaversli A, Hyde K. *Pestalotiopsis*-morphology, phylogeny, biochemistry and diversity. Fungal Di. 2011;50(1):167–87.
7. Han SH, Wang Y, Wang M, Li SH, Ruan R, Qiao T, et al. First report of *Pestalotiopsis microspora* causing leaf blight disease of *Machilus Nanmu* in China. Plant Dis. 2019;103(11):2963.
8. Chen Y, Wan Y, Zeng L, Meng Q, Yuan L, Tong H. Characterization of *Pestalotiopsis Chamaeropis* causing gray blight disease on tea leaves (*Camellia sinensis*) in Chongqing, China. Can J Plant Pathol. 2021;43(3):413–20.
9. González P, Alaniz S, Montelongo MJ, Rebellato R, Silvera-Pérez E, Mondino-González P, et al. First report of *Pestalotiopsis clavispor*a causing dieback on blueberry in Uruguay. Plant Dis. 2012;96(6):914.
10. Solarte F, Muñoz C, Maharachchikumbura S, ÁlvarezSolarte E. Diversity of *Neopestalotiopsis* and *Pestalotiopsis* Sp, causal agents of *Guava scab* in Colombia. Plant Dis. 2017;102(1):49–59.
11. Bhuiyan M, Islam S, Bukhari M, Kader M, Chowdhury M, Alam M, et al. First report of *Pestalotiopsis microspora* causing leaf blight of banana in Bangladesh. Plant Dis. 2021;106(5):1518.
12. Bhuiyan M, Islam S, Bukhari M, Kader M, Chowdhury M, Alam M, et al. First report of *Pestalotiopsis virgatula* causing *Pestalotiopsis* fruit rot on rambutan in Hawaii. Plant Dis. 2008;92(5):835.
13. Liu L. Bioactive metabolites from the plant endophyte *Pestalotiopsis Fici*. Mycology. 2011;2(1):37–45.
14. Klosterman S, Rollins J, Sudarshana M, Vinatzer B. Disease management in the genomics era-summaries of focus issue papers. Phytopathology. 2016;106(10):1068–70.
15. Wang X, Zhang X, Liu L, Xiang M, Wang W, Sun X, et al. Genomic and transcriptomic analysis of the endophytic fungus *Pestalotiopsis fici* reveals its lifestyle and high potential for synthesis of natural products. BMC Genomics. 2015;16(1):28.
16. Gibson DM, King B, Hayes M, BergstromGibson G. Plant pathogens as a source of diverse enzymes for lignocellulose digestion. Curr Opin Microbiol. 2011;14(3):264–70.
17. Drula E, Garron M, Dogan S, Lombard V, Henrissat B, Terraponet N, et al. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022;50(D1):D571–7.
18. Maharachchikumbura SSN, Hyde KD, Groenewald JZ, Xu J, Crous PW. Pestalotiopsis revisited. Stud Mycol. 2014;79(1):121–86.
19. Senanayake IC, Lian TT, Mai XM, Jeewon R, Maharachchikumbura SSN, Hyde KD, et al. New geographical records of *Neopestalotiopsis* and *Pestalotiopsis* species in Guangdong Province, China. Asian J Mycol. 2020;3(1):512–33.
20. Wang X, Dong H, Lan J, Liu Y, Liang K, Lu Q, et al. High-quality genome resource of the pathogen of *Diaporthe (Phomopsis) phragmitis* causing kiwifruit soft rot. Mol Plant Microbe in. 2020;34(2):218–21.
21. Wang G, Chen L, Tang W, Wang Y, Zhang Q, Wang H, et al. Identifying a melanogenesis-related candidate gene by a high-quality genome assembly and population diversity analysis in *Hypsizygus marmoreus*. J Genet Genomics. 2021;48(1):75–87.
22. Li L, Pan H, Liu W, Chen M, Zhong C. First report of *Diaporthe actinidiae* causing stem-end rot of kiwifruit during post-harvest in China. Plant Dis. 2017;101(6):1054.
23. Wicker T, Sabot F, Hua-Van A, Bennetzen J, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8(12):973–82.
24. Vergara Z, Sequeira-Mendes J, Morata J, Peiró R, Hénaff E, Costas C, et al. Retrotransposons are specified as DNA replication origins in the gene-poor regions of *Arabidopsis* heterochromatin. Nucleic Acids Res. 2017;45(14):8358–68.
25. Havecker E, Gao X, Voytas D. The diversity of LTR retrotransposons. Genome Biol. 2004;5(6):225.
26. Mandal P, Kazazian H. .SnapShot: vertebrate transposons. Cell. 2008;135(1):192.
27. Bellincampi D, Cervone F, Lionetti V. Plant cell wall dynamics and wall-related susceptibility in plant-pathogen interactions. Front Plant Sci. 2014;5:228.
28. Popper Z, Michel G, Hervé C, Domozych D, Willats W, Tuohy M, et al. Evolution and diversity of plant cell walls: from algae to flowering plants. Annu Rev Plant Bio. 2011;62(1):567–90.
29. Xie J, Wei J, Wang K, Luo J, Wu Y, Luo J, et al. Three phytotoxins produced by *Neopestalotiopsis Clavispora*, the causal agent of ring spot on *Kadsura coccinea*. Microbiol Res. 2020;238:126531.
30. Wenneker M, Pham K, Boekhoudt L, Boer F, Leeuwen P, Hollinger T, et al. First report of *Truncatella Angustata* causing postharvest rot on 'Topaz' apples in the Netherlands. Plant Dis. 2016;101(3):508.
31. Jayani R, Saxena S, Gupta R. Microbial pectinolytic enzymes: a review. Process Biochem. 2005;40(9):2931–44.
32. Yadav K, Dwivedi S, Gupta S, Tanveer A, Yadav S, Yadav P, et al. Recent insights into microbial pectin lyases: a review. Process Biochem. 2023;134:199–217.
33. Li S, Darwish O, Alkharouf N, Musungu B, Matthews B. Analysis of the genome sequence of *Phomopsis longicolla*: a fungal pathogen causing *Phomopsis* seed decay in soybean. BMC Genomics. 2017;18(1):688.
34. Zhao Z, Liu H, Wang C, Xu J. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. BMC Genomics. 2013;14(1):274.
35. Duplessis S, Cuomo C, Lin Y, Aerts A, Tisserant E, Veneault-Fourrey C, et al. Obligate biotrophy features unraveled by the genomic analysis of rust fungi. P Natl Acad Sci USA. 2011;108(22):9166–71.
36. Martin F, Aerts A, Ahrén D, Brun A, Danchin E, Duchaussoy F, et al. The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature. 2008;452(7183):88–92.
37. Chang H, Yendrek C, Caetano-Anolles G, Hartman G. Genomic characterization of plant cell wall degrading enzymes and in silico analysis of xylanses and polygalacturonases of *Fusarium Virguliforme*. BMC Microbiol. 2016;16(1):147.

38. Tariqjaveed M, Mateen A, Wang S, Qiu S, Zheng X, Zhang J, et al. Versatile effectors of phytopathogenic fungi target host immunity. J Integr Plant Biol. 2021;63(11):1856–73.
39. Ma Z, Song T, Zhu L, Ye W, Wang Y, Shao Y, Dong S, Zhang Z. A *Phytophthora sojae* glycoside hydrolase 12 protein is a major virulence factor during soybean infection and is recognized as a PAMP. Plant Cell. 2015;27(7):2057–72.
40. Qutob D, Kemmerling B, Brunner F, Küfner I, Engelhardt S, Gust A, et al. Phytotoxicity and innate immune responses induced by Nep1-like proteins. Plant Cell. 2006;18(12):3721–44.
41. Jonge R, Esse H, Kombrink A, Shinya T, Desaki Y, Bours R, et al. Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science. 2010;329(5994):953–55.
42. Bolger A, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
43. Koren S, Walenz B, Berlin K, Miller J, Bergman N, Phillippyet A. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36.
44. Chin C, Peluso P, Sedlazeck F, Nattestad M, Concepcion G, Clum A, et al. vPhased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016;13(12):1050–54.
45. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
46. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protocols Bioinf. 2009;25(1):4101–14.
47. Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
48. Hoff K, Stanke M. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. Nucleic Acids Res. 2013;41(W1):W123–8.
49. Stanke M, Diekhans M, Baertsch R, HausslerD. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24(5):637–44.
50. Simão F, Waterhouse R, Ioannidis P, Kriventseva E, Zdobnov E. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–12.
51. Krzywinski M, Schein K, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
52. Sayers E, Bolton E, Brister J, Canese K, Chan J, Comeau D, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50(D1):D20–6.
53. Bairoch A, Boeckmann B, Ferro S, Gasteiger E. Swiss-Prot: juggling between evolution and stability. Brief Bioinform. 2004;5(1):39–55.
54. Kent W. BLAT–the BLAST-like alignment tool. Genome Res. 2002;12(4):656–64.
55. Huerta-Cepas J, Forslund K, Coelho L, Szklarczyk D, Jensen L, Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34(8):2115–22.
56. Tatusov R, Fedorova N, Jackson J, Jacobs A, Kiryutin B, Koonin E, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics. 2003;4(1):41.
57. Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35(suppl2):W182–5.
58. Wang Y, Tang H, DeBarry J, Tan X, Li J, Wang J, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
59. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40(W1):W445–51.
60. Khaldi N, Seifuddin F, Turner G, Haft D, Nierman W, Wolfe K, et al. SMURF: genomic mapping of fungal secondary metabolite clusters. Fungal Genet Biol. 2010;47(9):736–41.
61. Medema M, Blin K, Cimermancic P, Jager V, Zakrzewski P, Fischbach M, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39(suppl 2):W339–46.

Publisher's note