




## Article

# The First Telomere-to-Telomere Chromosome-Level Genome Assembly of *Stagonospora tainanensis* Causing Sugarcane Leaf Blight

Fu Xu <sup>1,†</sup>, Xiuxiu Li <sup>2,†</sup>, Hui Ren <sup>1</sup>, Rensen Zeng <sup>1</sup> , Zhoutao Wang <sup>1</sup>, Hongli Hu <sup>2</sup>, Jiandong Bao <sup>3,\*</sup>   
and Youxiong Que <sup>1,\*</sup> 

<sup>1</sup> Key Lab of Sugarcane Biology and Genetic Breeding, Ministry of Agriculture and Rural Affairs, Fuzhou 350002, China

<sup>2</sup> College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou 350002, China

<sup>3</sup> State Key Laboratory for Managing Biotic and Chemical Treatments to the Safety of Agro-Products, Institute of Plant Protection and Microbiology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

\* Correspondence: baojd@zaas.ac.cn (J.B.); queyouxiong@126.com or queyouxiong@fafu.edu.cn (Y.Q.); Tel.: +86-591-8385-2547 (Y.Q.)

† These authors contributed equally to this work.

**Abstract:** The sexual morph *Leptosphaeria taiwanensis* Yen and Chi and its asexual morph *Stagonospora tainanensis* W. H. Hsieh is an important necrotrophic fungal phytopathogen, which causes sugarcane leaf blight, resulting in loss of cane tonnage and sucrose in susceptible sugarcane varieties. Decoding the genome and understanding of the basis of virulence is vitally important for devising effective disease control strategies. Here, we present a 38.25-Mb high-quality genome assembly of *S. tainanensis* strain StFZ01, denovo assembled with 10.19 Gb Nanopore sequencing long reads (~267×) and 3.82 Gb Illumina short reads (~100×). The genome assembly consists of 12 contigs with N50 of 2.86 Mb of which 5 belong to the telomere to telomere (T2T) chromosome. It contains 13.20% repeat sequences, 12,543 proteins, and 12,206 protein-coding genes with the BUSCO completeness 99.18% at fungi ( $n = 758$ ) and 99.87% at ascomycota ( $n = 1706$ ), indicating the high accuracy and completeness of our gene annotations. The virulence analysis in silico revealed the presence of 2379 PHIs, 599 CAZys, 248 membrane transport proteins, 191 cytochrome P450 enzymes, 609 putative secreted proteins, and 333 effectors in the StFZ01 genome. The genomic resources presented here will not only be helpful for development of specific molecular marker and diagnosis technique, population genetics, molecular taxonomy, and disease managements, it can also provide a significant precise genomic reference for investigating the ascomycetous genome, the necrotrophic lifestyle, and pathogenicity in the future.

**Keywords:** *Stagonospora tainanensis*; sugarcane leaf blight; pathogenicity; Nanopore sequencing; genome assembly



**Citation:** Xu, F.; Li, X.; Ren, H.; Zeng, R.; Wang, Z.; Hu, H.; Bao, J.; Que, Y. The First Telomere-to-Telomere Chromosome-Level Genome Assembly of *Stagonospora tainanensis* Causing Sugarcane Leaf Blight. *J. Fungi* **2022**, *8*, 1088. <https://doi.org/10.3390/jof8101088>

Academic Editors: Katrina Maria Ramonell and Yanbin Yin

Received: 18 September 2022

Accepted: 13 October 2022

Published: 16 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Sugarcane (*Saccharum* spp. hybrids), cultivated in more than 120 countries, is a crucial sugar crop accounting for 80% of the world's and nearly 90% of China's sugar production [1,2]. Similar to other crops, sugarcane is also exposed to many different diseases during cultivation. Among them, fungal diseases are the most serious due to the production of a large number of conidia, which can be transmitted by air, wind, and the splash of water during rain, and lead to the elimination of many elite cultivars [3–6]. Unlike the stalk-infected diseases, such as smut and pokkah boeng caused by *Sporisorium scitamineum* and *Fusarium* sp., respectively, which lead to serious economic losses almost every year in susceptible varieties, the foliar diseases are confined to leaves and outbreaks in susceptible varieties during monsoon, which increase wounds and humidity, a steady stream of wet weather and after the monsoon season. It means that the foliar diseases

do not cause outbreaks every year and thus the economic repercussions caused by them are uncertain. Therefore, less attention has been paid to the foliar diseases especially for those with limited distribution, such as sugarcane leaf blight (SLB). However, we note an increasing concern of foliar diseases in sugarcane due to minor diseases becoming major diseases. For example, pokkah boeng was a minor disease in past decades [7], but it has become a major disease in India [8,9]. The first severe outbreaks of brown rust caused by *Puccinia melanocephala* on sugarcane were reported in 1978 in Florida [10] and recently reported in India [11]. Orange rust caused by *Puccinia kuehnii* was also considered a minor disease in most countries, including Australian before 1999; however, severe epidemics occurred in cultivar Q124 during 1999–2001, which resulted in 50% of yield losses, overall losses estimated to be A\$150–210 millions [4], and an outbreak of this disease in America in 2007 [12]. Additionally, brown spot caused by *Cercospora longipes* was found to be alarming in India [13], peanut collar rot caused by *Aspergillus niger* in Asia [14,15], and brown stripe caused by *Helminthosporium stenospila* have been major diseases in China [16]. Similarly, SLB has been alarming recently in Yunnan and Guangxi provinces in China, due to the changed climate and the expansion of susceptible varieties, such as Guitang42, Taitang25, and Liucheng03-182 [16].

Sugarcane leaf blight caused by *Stagonospora tainanensis* was first reported in the year 1952 in Taiwan, China, and its asexual morph was first named *Cercospora tainanensis* [17] and then changed to *S. tainanensis* W. H. Hsieh, according to the further study on the isolates and inoculation, together with the pathogenic morphological features [18]. It occurred through the year in the east coast area with high rainfall, but not in the west coast area with less rain fall. The prevalence conditions for SLB were similar to that of the disease of *Stagonospora nodorum* blotch caused by *Stagonospora. nodorum* in wheat [19,20] of which the sexual morph is *Leptosphaeria nodorum* [21]. For a long time, research on SLB has been limited to pathogen isolation and identification, although it is one of the most harmful fungal diseases threatening the sugarcane industry and causing high cane yield and sugar losses in susceptible cultivars [17,22] because the pathogenic conidia are highly virulent, cause blight symptoms on sugarcane leaves, and result in loss of photosynthetic capacity [23]. Morphologically, *S. tainanensis* is an ascomycete, belonging to *Stagonospora* of Phaeosphaeriaceae within the class Dothideomycetes and the order Pleosporales. Its asci are slightly curved, 62–115 × 21–33 µm in size, born in scattered and dark brown perithecia [22]. Each ascus contains eight fusiform bent ascospores with one septum and its cylindrical conidia are straight to slightly curved in shape with three septa generally, containing three to eight oil droplets [20]. Recently, the efficient PCR detection technology of *S. tainanensis* and SLB were developed based on genomic information [24], and identification of SLB-resistance associated loci/genes were reported [25,26]. These will advance our understanding of the molecular mechanisms of pathogen infection and host resistance.

In the recent 10 years, the revolutionary progress of the third-generation sequencing (TGS) technologies led by PacBio and Oxford Nanopore Technology (ONT) has brought genome research into a new era [27,28]. Third-generation sequencing can produce long reads from 10 kb to 1 Mb, which dramatically reduce the time and cost for genome assembling and thus makes it possible to finish a high-quality non-model fungus genome assembly of approximately 50 Mb for a small lab [29]. For a model plant pathogenic fungi, such as *Pyricularia oryzae* and *Fusarium graminearum*, TGS-based chromosome-level reference genome and more than 100 genome assemblies are available in NCBI [30–33]. For fungi in the *Massarinaceae* family, only three species, including *Bysotheций circinans* [34], *Massarina eburnea* [35], and *Stagonospora* sp. [36], have been reported. However, in our studied genus *Stagonospora*, only *S. nodorum*, a model species of necrotrophic Pleosporales pathogens, had been sequenced. Additionally, *S. nodorum* genome was the first Dothideomycete genome, which was sequenced in 2005 and published in 2007 [37], and had a revolutionary impact on the understanding of this important pathogen and other fungal pathogens due to the limited genomic information available.

Until now, the genome of *S. tainanensis* has not been publicly available. In this study, we employed Nanopore sequencing and Illumina sequencing together to finish a near telomere-to-telomere chromosome-level genome assembly and RNA-seq based gene annotation of this necrotrophic infecting fungus *S. tainanensis* strain StFZ01. It can provide a more precise understanding of the pathogen and the fungal pathogenicity and can offer a series of putative proteins in the fungal pathogenesis, such as effectors, and it is thus beneficial for developing a new disease management strategy and for sugarcane improvement of leaf blight resistance.

## 2. Materials and Methods

### 2.1. Sample Preparation and Sequencing

Sugarcane leaf blight-susceptible sugarcane cultivar Yuetang93-159 was planted in Fuzhou, China (26°5'0" N, 119°13'45" E) and the *S. tainanensis* strain StFZ01 was collected from its leaves with the typical symptoms of SLB. After the isolates were testified using ITS detection (Figure 1) and by pathogenic morphology [24], one assigned as StFZ01 was used for genome sequencing and analysis. The fresh mycelia cultivated on potato dextrose broth (PDB) media was collected for DNA and RNA extraction. For long-read genomic sequencing, high-quality genomic DNA was extracted using Ligation Sequencing Kit (SQK-LSK110), then BluePippin DNA size selection system was used to select large DNA fragments (>20 kb) for sequence library preparation following the manufacturer's instructions, and the sequencing was conducted on *PromethION* sequencing platform from Oxford Nanopore Technologies (ONT). For Illumina short-read sequencing, genomic DNA and mRNA were extracted, purified, and prepared for sequencing libraries using Illumina DNA Prep Kits (Illumina, Inc., San Diego, CA, USA) and Illumina Stranded mRNA Prep (Illumina, Inc., San Diego, CA, USA), respectively. Illumina genomic DNA sequencing and RNA-seq were performed on the Illumina HiSeq 3000 sequencing platform (350 bp library and PE150 strategy).

### 2.2. Genome Size Estimation

The genome size of strain StFZ01 was estimated using GenomeScope v2.0 [38] based on the k-mer frequencies of genomic Illumina short reads (k = 21, p = 1) computed by KMC v3.1.1 [39].

### 2.3. De Novo Genome Assembly

The de novo assembly of ONT long reads was performed using NextDenovo v2.5.0 (<https://github.com/Nextomics/NextDenovo>) (accessed on 10 January 2022) with the "correct-then-assemble" strategy. Then, base errors (SNPs/Indels) of the draft genome assembly were fixed by NextPolish v1.4.0 [40] using both ONT long reads and Illumina short reads (task = best model) to generate a high-continuity and high-accuracy genome assembly of strain StFZ01.

### 2.4. Genome Completeness Assessment

The software Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.3.2 [41] was used to evaluate the completeness of the genome assembly and annotated genes with the lineage dataset of fungi\_odb10 ( $n = 758$ ) and ascomycota\_odb10 ( $n = 1706$ ). Furthermore, the completeness of the genome assembly was also assessed by mapping of sequenced reads. RNA-Seq reads were aligned to the repeat-masked genome assembly using HISAT2 v2.2.1 [42]. The ONT long reads and genomic Illumina short reads were mapped to the unmasked genome assembly with minimap2 v2.21-r1071 [43] and BWA-MEM2 v2.2.1 [44], respectively (Table S1).

Telomeric repeats (5'-TTAGGG-3') $n$  or (5'-CCCTAA-3') $n$  were searched at both ends of the contigs. The gapless contig ending with telomeric repeats should be an telomere-to-telomere (T2T) chromosome [45].



gous protein sequences in fungi\_odb10 (<https://busco-data.ezlab.org/v5/data/lineages/>) (accessed on 31 December 2021) as well as a de novo transcriptome assembly generated from RNA-seq data in this study.

### 2.7. Identification of Non-Coding RNAs

Transfer RNAs (tRNAs) were identified by tRNAscan-SE v2.0.9 with eukaryote parameters [50]. Ribosomal RNA (rRNA) and other non-coding RNAs (ncRNAs) were predicted using Infernal v1.1.4 [51] by searching against the RNA families database Rfam v14.8 [52].

### 2.8. Functional Annotation of Protein-Coding Genes

#### 2.8.1. General Functional Annotation

Gene Ontology (GO) and Pfam terms were grouped into categories based on results from InterProScan v5.55-88.0 [53]. Eukaryotic Orthologous Groups (KOG) were annotated by eggNOG-mapper v2 [54]. The predicted protein-encoding genes were assigned KEGG (Kyoto Encyclopedia of Genes and Genome) using KofamKOALA [55], then these KOs were used to reconstruct KEGG metabolic pathway maps at KEGG Mapper Reconstruction online web service ([https://www.genome.jp/kegg-bin/find\\_pathway\\_object](https://www.genome.jp/kegg-bin/find_pathway_object)) (accessed on 8 June 2022).

#### 2.8.2. Fungal Pathogenicity-Related Gene Annotation

Annotation of pathogenicity-related genes was conducted by DIAMOND v2.0.11 [56] against a set of databases including PHI-base v4.12 (<http://www.phi-base.org/>) (accessed on 21 September 2021), dbCAN2 (<https://ccb.unl.edu/dbCAN2>) (accessed on 31 July 2022), and TCDBs (<https://tcdb.org/>) (accessed on 31 July 2022) for identification of pathogen–host interaction related genes (similarity  $\geq 30\%$  and alignment length  $\geq 100$  aa), carbohydrate-active enzymes (CAZys) (verified by HMMER v3.1b2 [57] via bCAN and dbCAN-sub databases, number of tools  $\geq 2$ ), and membrane transport proteins (similarity  $\geq 50\%$  and alignment length  $\geq 200$  aa), respectively.

The putative secreted proteins were identified following a pipeline in the previous study [30], of which proteins with signal peptide and without transmembrane helix were identified by SignalP v5.0 [58] and TMHMM v2.0 [59], respectively, and those with extra-cellular location were identified using ProtComp v9.0 from MolQuest v2.4 (Softberry Inc., New York, USA). Furthermore, these effectors were further scanned by EffectorP v3.0 [60] and divided into cytoplasmic and apoplast effectors.

#### 2.8.3. Secondary Metabolite Biosynthetic Gene Clusters Analysis

Online web service (<https://fungismash.secondarymetabolites.org>) (accessed on 8 June 2022) of AntiSMASH v6.0 [61] for fungi was employed for identification of secondary metabolite biosynthetic gene clusters (SMBGCs).

### 2.9. Comparative Genomic Analysis

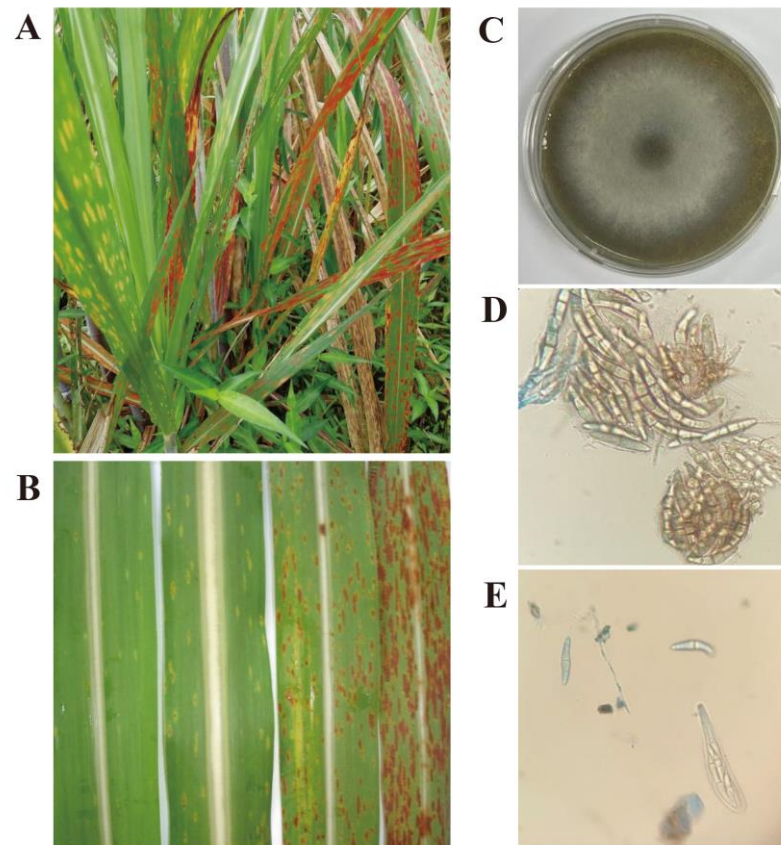
Whole-genome protein sequences of nine fungi species, including three of the *Massariaceae* family and six of the *Pleosporaceae* family in the *Pleosporales* order, were downloaded from NCBI (Table S2). The longest proteins for each gene were selected and clustered using OrthoFinder v2.5.4 [51] with the following parameters: -S diamond, -M msa [62]. Single-copy core orthologous proteins were aligned using MAFFT v7.490 [63] and then the phylogenetic tree of species was constructed with FastTree v2.1.11 [64] and visualize by Interactive Tree Of Life (iTOL) v6.5.8 online services [65].

## 3. Results and Discussion

### 3.1. The Morphology of the Pathological Lesions and Pathogenic *S. tainanensis* Used for Genome Sequencing

The typical single mature or early-mature lesion of sugarcane leaf blight on the infected leaves caused by *S. tainanensis* is spindly and elongated, which is observed on variety

Yuetang93-159 (Figure 2A). The color of pathological leaf tissues was found to change from early yellowish to yellow, then to red-brown (Figure 2B). Therefore, the typical symptoms of SLB are relatively easy to identify in later stages of lesion development (Figure 2B). The mycelia of the pathogenic *S. tainanensis* strain StFZ01 used for DNA isolating were collected from the culture growth on PDA agar medium (Figure 2C) and the pathogenic conidia (Figure 2D), and sexual asci and ascospores (Figure 2E) were observed.



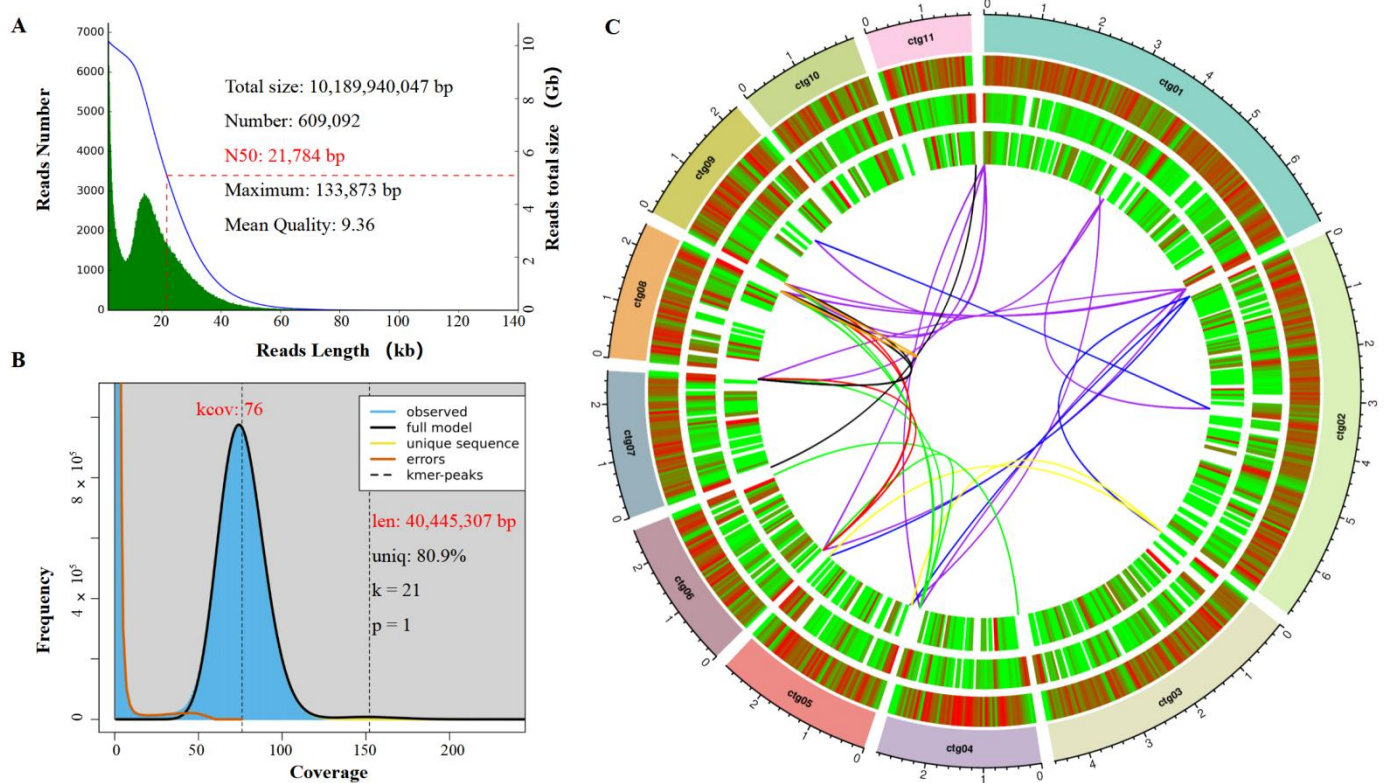
**Figure 2.** The phenotypic symptoms of sugarcane leaf blight (SLB) on sugarcane cultivar Yuetang93-159 initiated by *Stagonospora tainanensis* and the morphological characteristics of the strain StFZ01. (A) The phenotypic symptoms of SLB on plants. (B) Development of the lesions of SLB on leaves. (C) The pathogenic colony and mycelia. (D) The pathogenic conidia. (E) The sexual asci and ascospores.

### 3.2. Genome Sequencing and Assembly

After quality control, a total of 10.19 Gb clean ONT long reads (depth:  $\sim 267\times$ ,  $N_{50}$ : 21,784 bp, maximum length: 133,873 bp) were used for de novo genome assembly, 3.82 Gb Illumina short reads (depth:  $\sim 100\times$ ,  $2 \times 150$  bp) for estimation of genome size and polishing of draft genome assembly, and 6.08 Gb RNA-seq reads ( $2 \times 150$  bp) for gene annotation (Figure 3A and Table S1).

We estimated genome size of *S. tainanensis* strain StFZ01 to be 40,445,307 bp (Model fit = 98.13%, ploidy = haploid), based on k-mer distribution ( $k = 21$ , average k-mer depth  $76\times$ ) of Illumina short reads (Figure 3B and Table S3). The estimated repeat is 19.09% (7,720,991 bp) of the genome sequence. The 38.25 Mb genome assembly (GC: 51.49%) of StFZ01 contained 12 contigs with  $N_{50}$  of 2.86 Mb ( $L_{50} = 4$ ),  $N_{90}$  of 2.11 Mb ( $L_{90} = 10$ ), and a maximum contig of 7.12 Mb (Table 1 and Figure 3C). The length of genome assembly is slightly smaller than the estimated genome size (94.58% of 40,445,307 bp). Its genome size is comparable to another species *S. nodorum* (37.21 Mb) in the same genus and *Bipolaris maydis* (36.23 Mb) causing Southern corn leaf blight [66], but smaller than the *Leptosphaeria*

*maculans* (45.12 Mb), a pathogenic fungus closely related to *S. nodorum*, and much smaller than the genomes of *Colletotrichum higginsianum* (53.4 Mb) and *Colletotrichum graminicola* (57.4 Mb) [67].



**Figure 3.** Genome features of *Stagonospora tainanensis* strain StfZ01. (A) Summary of ONT reads. (B) Genome size estimation with NGS genomic reads; (C) Circos plot of genome assembly features. Circles from outside to inside present contigs (1st circle, the smallest contig ctg12 was not shown), distribution of protein-coding genes (2nd), TEs (3rd), and putative secreted proteins (4th) per 50 kb window size (color blue to red means number from low to high). The lines in the center of circle show the synteny blocks ( $\geq 10$  kb) between different contigs.

**Table 1.** Genome Assembly Features of *Stagonospora Tainanensis* Strain StfZ01.

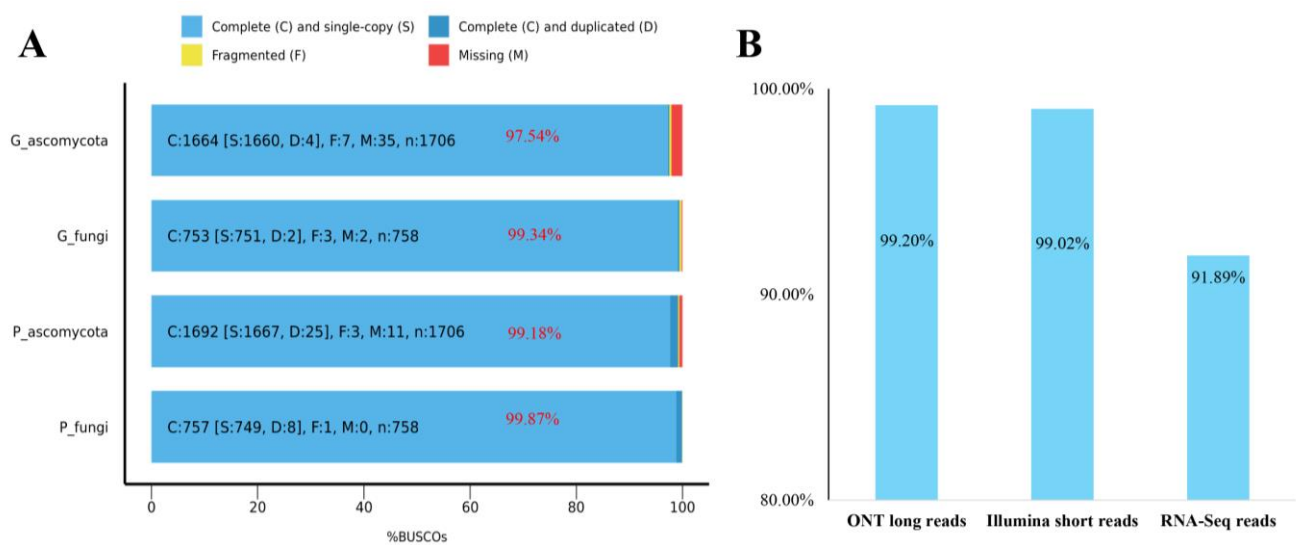
Features	StfZ01
Assembly size (bp)	38,252,541
Contig number	12
Contig N50 (bp)	2,858,663
L50	4
Contig N90 (bp)	2,113,312
L90	10
Average contig length (bp)	3,187,712
Maximum contig length (bp)	7,120,155
GC content	51.49%
Repeat sequences	13.20%
Protein-coding genes	12,206
tRNAs	162
rRNAs	142
Other ncRNA	43

A total of nine contigs (ctg1-ctg9) were found to start or end with telomeric repeat, (5'-TTAGGG-3')<sub>n</sub> or (5'-CCCTAA-3')<sub>n</sub>, of which five contigs (ctg1, ctg3, ctg5, ctg6, ctg7, and

ctg9) contain telomeric repeats at both contig ends, indicating that these contigs reached perfect gapless T2T chromosome level [45] (Table S4).

### 3.3. Genome Quality Assessment

The BUSCO completeness values were estimated to be 99.34% at fungi ( $n = 758$ ) and 97.54% at ascomycota ( $n = 1706$ ) for genome assembly of StFZ01 (Figure 4A). All clean ONT long reads and Illumina genomic reads were aligned to unmasked genome assembly, and the mapping rate of ONT long reads and Illumina genomic reads are 99.20% and 99.02%, respectively (Figure 4B and Table S1). Furthermore, one RNA-seq sample of StFZ01 was mapped to repeat-masked genome assembly, and 91.89% (85.9% properly paired) of RNA-seq reads showed unique mapping to gene regions (Figure 4B and Table S1). All of these results attest to the high continuity and completeness of our assembled genome.



**Figure 4.** Genome completeness assessment of *Stagonospora tainanensis* strain StFZ01. **(A)** BUSCO completeness assessment of genome assembly and annotated proteins with the lineage dataset of fungi\_odb10 and ascomycota\_odb10. G means genome and P means proteins. **(B)** Genome completeness valued by mapping rate of different reads.

### 3.4. Repeat Analysis

Repetitive sequences were identified using a combination of ab initio and homology-based approaches. In total, 13.20% (5,048,126 bp) of the assembled StFZ01 sequences were annotated as repeat sequences (Tables 1 and 2), which is less than the *C. graminicola* (22.3%) but more than the *C. higginsianum* (9.1%) [67]. Interspersed repeats, as the major component (90.41% of total repeats) were found to account for 11.28% of the genome, including 1,927,917 bp long terminal repeats (LTRs), 1,242,571 bp DNA transposons, 498,662 bp long interspersed nuclear elements (LINEs), 12,515 bp short interspersed nuclear elements (SINEs), and 1,242,571 bp unclassified interspersed repeats (Table 2). The dynamic polymorphism of repeat insertion in phytopathogenic fungi usually associated with virulence variations, hence, high frequent interspersed repeats will be candidate DNA markers for identification of different virulent strain, similar to Pot2 rep-PCR fingerprinting analysis in rice blast fungus [68]. In addition, we identified several types of non-coding RNAs, including 162 tRNAs, 142 rRNAs, and 48 other ncRNAs (Table 1).



**Table 2.** Repeats Identified in *Stagonospora Tainanensis* Strain StFZ01.

	Features	Count	Length (bp)	Percentage (%)
Interspersed repeats	SINEs <sup>1</sup>	62	12,515	0.03
	LINEs <sup>2</sup>	288	498,662	1.30
	LTR <sup>3</sup> elements	2592	1,927,917	5.04
	DNA transposons	1101	1,242,571	3.25
	Unclassified	3246	882,135	2.31
Tandem repeats	Small RNA	137	111,320	0.29
	Simple repeats	7496	312,397	0.82
	Low complexity	1175	60,609	0.16
	Total repeats	-	5,048,126	13.20

<sup>1</sup> short interspersed nuclear elements; <sup>2</sup> long interspersed nuclear elements; <sup>3</sup> long terminal repeat.

### 3.5. Gene Structural Annotation

In total, 12,206 high-confidence protein-coding genes were predicted by the BRAKER2 pipeline, which was more than the *Magnaporthe grisea* PMg\_DI (10,218) though its genome size (38.25 Mb) was less than the *M. grisea* (47.89 Mb) [69], and this phenomenon was also observed in the other necrotrophic fungal pathogen *Pyrenophora teres f. teres* (41.95 Mb size, 11,799 genes) [70] and *L. maculans* 'brassicae' WA74 (44.20 Mb size, 10,624 genes) [71]. However, the gene number and the genome size of *S. tainanensis* were comparable to *M. oryzae* (on average 12,684 genes, 40.12 Mb size) [72]. In addition, this number is less than the other species in the same genus *Stagonospora*, i.e., model species *S. nodorum* SN15 (37.02 Mb size, 17,580 genes) ([https://www.ncbi.nlm.nih.gov/assembly/GCA\\_016801405.1/](https://www.ncbi.nlm.nih.gov/assembly/GCA_016801405.1/)) (accessed on 30 August 2022), substantially higher than the known filamentous fungi though the gene number changed from first 10,792 supported by EST [35] to 12,382 supported by integrated multidimensional omics [73] and now 17,580 [74], while the gene number was comparable to the other typical species in filamentous fungi. These genes encode 12,543 proteins and 97.42% (11,891) of genes were predicted, encoding only one protein isoform. Only 315 genes were predicted with alternatively spliced protein isoforms, including 294 genes encoding two protein isoforms, 20 genes encoding three protein isoforms, and only 1 gene encoding 4 protein isoforms (Table S5). The number of exons per gene ranged from 1–18, and most genes contained 1–5 exons (Table S6). The BUSCO completeness of genes is 99.18% at fungi ( $n = 758$ ) and 99.87% at ascomycota ( $n = 1706$ ) (Figure 4A), indicating high accuracy and completeness of our gene annotations.

The gene distribution is not uniform on the contigs. In most cases, it is opposite to the distribution of repetitive sequences, especially in the repeat-rich telomeric region at both ends of the contig, which contains almost no genes (Figure 3C). This phenomenon is common in repeat-rich pathogenic fungi, such as like rice blast fungus *P. oryzae* [30] and soil borne plant pathogen *Verticillium dahliae* [75].

In addition, no protein-coding gene was identified in the smallest contig ctg12 (154,536 bp). Blastn against with NCBI nr database revealed it is mitochondrion DNA, the most similar sequence is mitochondrion from *B. sorokiniana* (NC\_047242.1, 92.83% similarity and 26% coverage). Thus, we did not show it in the genome circos plot (Figure 3C).

### 3.6. General Gene Functional Annotation

W General gene functional annotation estimated that 69.24% (8452) of genes contain conserved protein domains, and 41.95% (5121) were classified by GO terms, with 36.40% (4443) mapping to KEGG pathways and 74.37% (9078) assigned to KOG categories (Tables 3 and S7–S10).

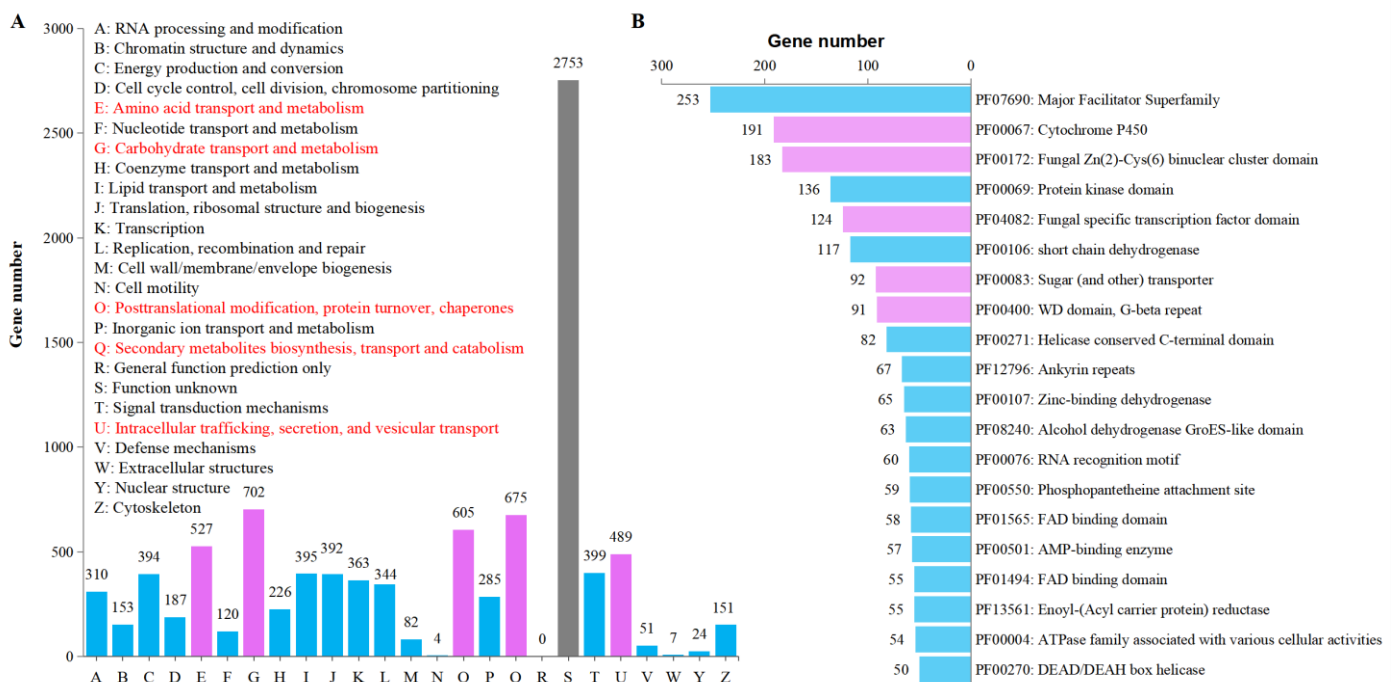
**Table 3.** Summary of Gene Functional Annotation.

Annotation	Gene Number	Percentage (%)
Pfam	8452	69.24%
GO <sup>1</sup>	5121	41.95%
KEGG <sup>2</sup>	4443	36.40%
KOG <sup>3</sup>	9078	74.37%
CAZys <sup>4</sup>	599	4.91%
PHIs <sup>5</sup>	2379	19.49%
Cytochrome P450 enzymes	191	1.56%
Membrane transport proteins	248	2.03%
Putative secreted proteins	606	4.96%
Effectors	332	2.72%
SMBGCs <sup>6</sup>	58	0.48%

<sup>1</sup> Gene Ontology; <sup>2</sup> Kyoto Encyclopedia of Genes and Genome; <sup>3</sup> EuKaryotic Orthologous Groups; <sup>4</sup> Carbohydrate-Active enZymes; <sup>5</sup> Pathogen–Host Interaction genes; <sup>6</sup> Secondary Metabolite Biosynthetic Gene Clusters.

The top five GO annotations mainly referred to protein transport and binding activity, including ‘transmembrane transport’ (562 genes), ‘transmembrane transporter activity’ (417 genes), ‘ATP binding’ (466 genes), ‘protein binding’ (446 genes), and ‘zinc ion binding’ (313 genes) (Table S11). The top KEGG annotation referred biosynthesis pathway, including ‘biosynthesis of secondary metabolites’ (325 genes), ‘microbial metabolism in diverse environments’ (179 genes), ‘biosynthesis of cofactors’ (113 genes), and ‘biosynthesis of amino acids’ (105 genes) (Table S12).

KOG annotated genes were divided into 24 categories of which, ‘E: Amino acid transport and metabolism’ (527 genes), ‘G: Carbohydrate transport and metabolism’ (702 genes), ‘O: Posttranslational modification, protein turnover, chaperones’ (605 genes), ‘Q: Secondary metabolites biosynthesis, transport and catabolism’ (675 genes), and ‘U: Intracellular trafficking, secretion, and vesicular transport’ (489 genes) were top five terms associated with candidate pathogenicity-related genes (Figure 5A).

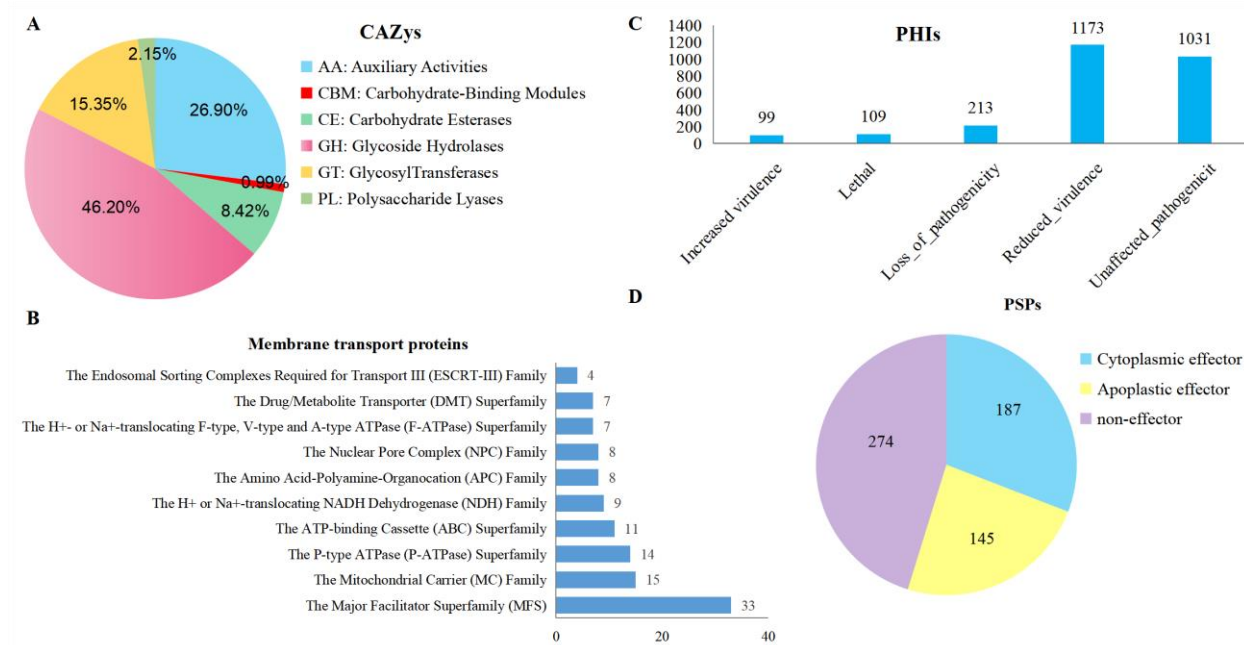


**Figure 5.** General gene functional annotation. (A) KOG annotation and (B) Pfam annotation (top20). Red color indicates the top terms associated with candidate pathogenicity-related genes.

The Pfam annotation revealed a set of candidate pathogenicity-related genes, including ‘Cytochrome P450’ (191 genes) (Table S13), ‘fungal Zn(2)-Cys(6) binuclear cluster domain’ (183 genes), ‘fungal specific transcription factor domain’ (124 genes), ‘Sugar (and other) transporter’ (92 genes), and ‘WD domain, G-beta repeat’ (91 genes) (Figure 5B).

### 3.7. Annotation of Pathogenicity-Related Genes

To understand the pathogenicity mechanism of *S. tainanensis*, we identified a large number of pathogenicity-related proteins, including 599 CAZys, 248 membrane transport proteins, and 2379 PHIs (Table 3), and the number of CAZys was more than in the *M. grisea* (539) [59]. CAZys mainly consisted of 280 (46.20%) glycoside hydrolases (GHs), 163 (26.90%) auxiliary activities (AAs), 93 (15.35%) glycosyl transferases (GTs), and 51 (8.42%) carbohydrate esterases (CEs) (Figure 6A and Table S14). The top three membrane transport proteins included 33 ‘The Major Facilitator Superfamily (MFS)’, 15 ‘The Mitochondrial Carrier (MC) Family’, 14 ‘The P-type ATPase (P-ATPase) Superfamily’, 11 ‘The ATP-binding Cassette (ABC) Superfamily’, and 9 ‘The H<sup>+</sup> or Na<sup>+</sup>-translocating NADH Dehydrogenase (NDH) Family’ (Figure 6B and Table S15).



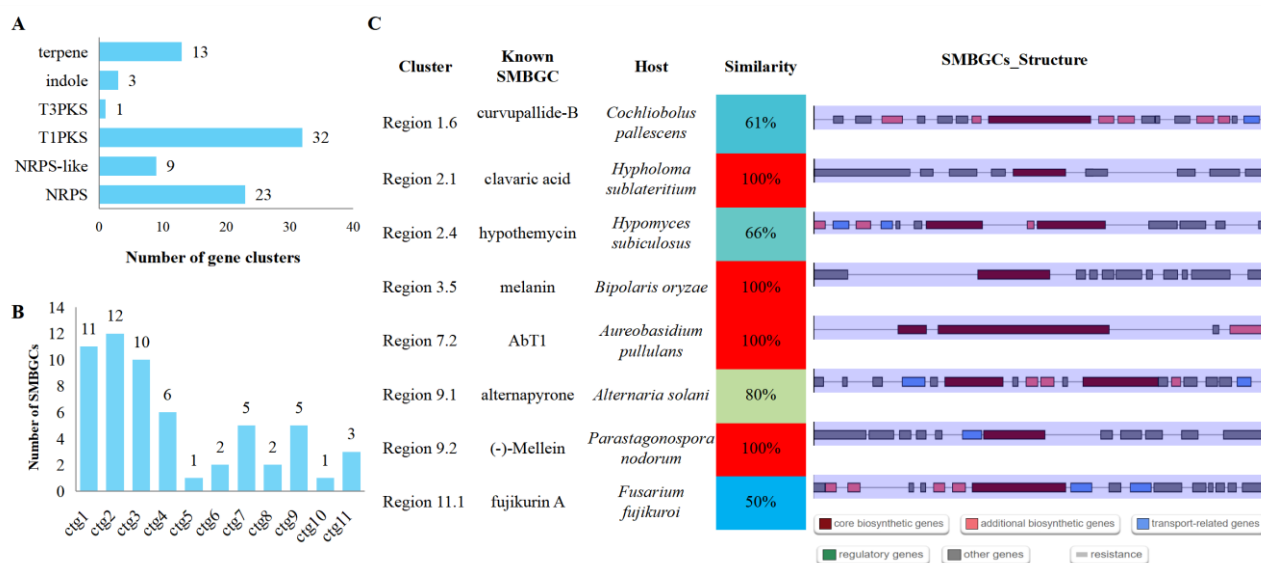
**Figure 6.** Summary of pathogenicity-related gene annotations. (A) CAZys, (B) Membrane transport proteins (top10), (C) PHIs, and (D) Putative secreted proteins.

Blast analysis of genomic loci using PHI-base revealed a total of 2625 PHI-associated genes, which is much more than the *M. grisea* PMg-ID (868) [74] but less than those in 16 *Sporothrix* strains (from 3083 to 4750) [76]. Among 2625 genes, 99 enhance virulence, 1173 reduce virulence, 109 are lethal, 213 are pathogenicity lost, and 1031 are pathogenicity unaffected (Figure 6C and Table S16).

We identified 1323 proteins with a signal peptide and 609 proteins (encoded by 606 genes) with extracellular location that were defined as putative secreted proteins (PSPs) after removing proteins containing transmembrane helix (Table 3). Among these secreted proteins, we deciphered 333 effectors (encoded by 332 genes), including 187 cytoplasmic and 145 apoplasmic effectors (Figure 6D and Table S17), and the number of effectors was much less than that in the *M. grisea* PMg-ID (594) [70]. Interestingly, we found the putative secreted proteins were enriched in or nearby the high repeat regions (Figure 3C), which is consistent with other pathogenic fungi, such as *Magnaporthe oryzae* [46].

### 3.8. Secondary Metabolite Biosynthetic Gene Clusters (SMBGCs)

A total of 58 SMBGCs were identified, including 23 nonribosomal peptide synthetases (NRPS), 9 NRPS-like, 32 type I polyketide synthases (T1PKS), 1 type III polyketide synthases (T3PKS), 3 indoles, and 13 terpenes (Figure 7A and Table S18). More than half of the SMBGCs (33) were located in ctg1, ctg2, and ctg3, the top three longest contigs (Figure 7B). Eight SMBGCs were found to have more than a 50% similarity with the known SMBGCs, with four having a 100% similarity with the clavatic acid biosynthetic gene cluster from *Hypholoma sublateritium*, melanin biosynthetic gene cluster from *B. oryzae*, AbT1 biosynthetic gene cluster from *Aureobasidium pullulans*, and (-)-Mellein biosynthetic gene cluster from *Parastagonospora nodorum*, respectively (Figure 7C). The rest of the 50 novel SMBGCs will provide a chance for mining novel secondary metabolites in *S. tainanensis*.



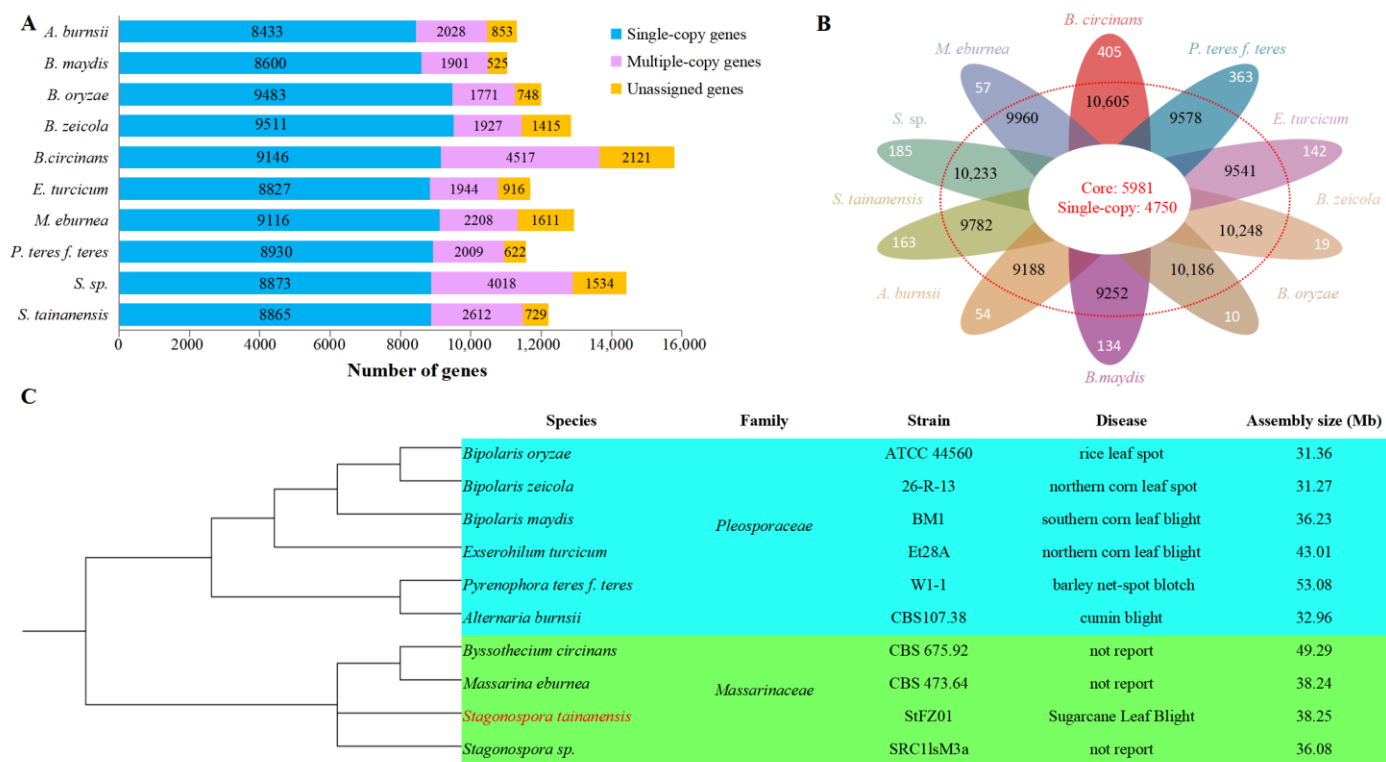
**Figure 7.** Profiles of secondary metabolite biosynthetic gene clusters. (A) Type of SMBGCs. (B) Contig distribution of SMBGCs. (C) Most similarity with known SMBGCs (similarity  $\geq$  50%).

Interestingly, clavatic acid was reported to be an inhibitor of the human Ras-farnesyl transferase [77,78]; it thus has antitumor activity [79], and its terpene biosynthetic gene cluster was also detected in *Aspergillus terreus* [80] and *Sporothrix* species [76]. Melanin, a black pigment synthesized by T1PKS type SMBGCs, has a central role in the pathogenicity of plant pathogenic fungi, such as rice blast fungus *P. oryzae* [81,82].

### 3.9. Comparative Genomic Analysis

To figure out the genome differentiation with close relationship species, whole-genome orthologous gene cluster analysis was performed among *S. tainanensis*, three species were from the *Massarinaceae* family and the other six species were from the *Pleosporaceae* family in the *Pleosporales* order (Table S2). We collected 125,793 genes from the 10 species, and 114,719 genes (91.20%) were clustered into 14,038 orthogroups (Table S19). For the genome of *S. tainanensis* StFZ01, 114,77 out of 12,206 genes (94.03%) were clustered into 9782 orthogroups, which include 5981 (61.14%) core orthogroups shared by all other nine fungi, and 47 (0.48%) species-specific orthogroups with 163 genes (Figure 8A,B and Table S20). Together with the 729 unclustered genes, we identified 892 species-specific genes (7.31%) in *S. tainanensis* StFZ01 (Figure 8A and Table S20).

We selected 4750 single-copy orthogroups to construct phylogenetic trees (Figure 8C), which showed that *S. tainanensis* and *Stagonospora sp.* were placed on a branch outside of *Massarina eburnea* and *Byssothecium circinans* (Figure 8C).



**Figure 8.** Comparative genomics analysis with close species. (A) Genes clustered in orthogroups. (B) Summary of orthogroups among species. Species specific genes (white), total orthogroups (black, in the red dotted circle), and core orthogroups (red) were shown from outside to inside of flower. (C) Phylogenetic tree inferred with alignment of single-copy core orthogroups.

### 4. Conclusions

In conclusion, this study presented the first T2T chromosome-level genome assembly and high-quality gene annotation of the pathogenic fungus *S. tainanensis* strain StFZ01 causing sugarcane leaf blight, integrating with Nanopore sequencing and Illumina sequencing. The well annotated repeats and genes, such as CAZys and effectors will play as the reference genome for designing species-specific molecular markers and identifying pathogenicity-related genes in the future.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jof8101088/s1>, Table S1: Summary of sequencing reads; Table S2: Whole-genome proteins date of close species used for orthologous gene clusters; Table S3: Genome size estimated by GenomeScope2 with Illumina genomic reads; Table S4: Contigs with telomeric repeats; Table S5: Distribution of protein isoforms encodes by genes; Table S6: Distribution of exon numbers per gene; Table S7: GO annotation; Table S8: KEGG annotation; Table S9: Pfam Annotation; Table S10: KOG annotation; Table S11: Summary of GO annotation; Table S12: Summary of KEGG pathway annotation; Table S13: Genes with cytochrome P450 annotation; Table S14: CAZys annotation; Table S15: Membrane transport proteins annotated by TCDB; Table S16: PHI\_annotation; Table S17: Summary of secreted proteins and effectors; Table S18: Annotation of secondary metabolite biosynthetic gene clusters by AntiSMASH; Table S19: Summary of orthogroups analysis overall species; Table S20: Summary of orthogroups analysis per species.

**Author Contributions:** Conceptualization, F.X., H.R., J.B. and Y.Q.; methodology, F.X., X.L., H.R., R.Z. and J.B.; software, F.X., X.L., H.H., Z.W. and J.B.; validation, F.X., X.L., Z.W. and J.B.; writing—original draft preparation, F.X., X.L., H.R., H.H. and J.B.; writing—review and editing, X.L., J.B., R.Z. and Y.Q.; supervision, R.Z. and Y.Q.; project administration, J.B. and Y.Q. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the fund earmarked for the Modern Agriculture Technology of China (CARS-17) and the Special Fund for Science and Technology Innovation of Fujian Agriculture and Forestry University (CXZX2020083A).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The whole genome sequence (GWHBJVM00000000) and the raw sequence data (CRA007499) reported in this paper have been respectively deposited in the WGH (Genome Warehouse, <https://ngdc.cncb.ac.cn/gwh>) and the GSA (Genome Sequence Archive, <https://ngdc.cncb.ac.cn/gsa>) in National Genomics Data Center (NGDC), China National Center for Bioinformatics (CNCB) [83], under Bioproject PRJCA010492.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Li, Y.-R.; Yang, L.-T. Sugarcane agriculture and sugar industry in China. *Sugar Tech* **2015**, *17*, 1–8. [CrossRef]
- Xu, F.; Wang, Z.T.; Lu, G.L.; Zeng, R.S.; Que, Y.X. Sugarcane ratooning ability: Research status, shortcomings, and prospects. *Biology* **2021**, *10*, 1052. [CrossRef] [PubMed]
- Raid, R.N. Physiological specialization in sugarcane rust (*Puccinia melanocephala*) in Florida. *Plant Dis.* **1989**, *73*, 183. [CrossRef]
- Rajput, M.A.; Rajput, N.A.; Syed, R.N.; Lodhi, A.M.; Que, Y. Sugarcane smut: Current knowledge and the way forward for management. *J. Fungi* **2021**, *7*, 1095. [CrossRef]
- Hoy, J.W.; Hollier, C.A. Effect of brown rust on yield of sugarcane in Louisiana. *Plant Dis.* **2009**, *93*, 1171–1174. [CrossRef]
- Shan, H.; Li, W.; Zhang, R.; Wang, X.; Li, J.; Cang, X.; Yin, J.; Luo, Z.; Huang, Y. Analysis on epidemic reason of sugarcane pokkah boeng and its losses on yield and sucrose content. *Sugar Crop China* **2018**, *40*, 40–42, 45.
- Patil, A.S.; Hapase, D.G. Studies on pokkah boeng disease of sugarcane in Maharashtra. *Indian Phytopathol.* **1987**, *40*, 290.
- Vishwakarma, S.K.; Kumar, P.; Nigam, A.; Singh, A.; Kumar, A. Pokkah boeng: An emerging disease of sugarcane. *J. Plant Pathol. Microb.* **2013**, *4*, 1000170.
- Viswanathan, R. Changing scenario of sugarcane diseases in India since introduction of hybrid cane varieties: Path travelled for a century. *J. Sugarcane Res.* **2018**, *8*, 1–35.
- Comstock, J.C.; Glynn, N.C.; Davidson, R.W. Sugarcane rusts in Florida. *Proc. Inter. Soc. Sugar Cane Tech.* **2010**, *27*, 1–9.
- Selvakumar, R.; Viswanathan, R. History of sugarcane rusts in India. In *Res Accomplishments in Sugarcane Pathology*; Viswanathan, R., Sundar, R.A., Malathi, P., Selvakumar, R., Jayakumar, V., Nithya, K., Eds.; ICAR-Sugarcane Breeding Institute: Coimbatore, India, 2018; pp. 76–79.
- Comstock, J.C.; Sood, S.G.; Glynn, N.C.; Shine, J.M., Jr.; McKemy, J.M.; Castlebury, L.A. First report of *Puccinia kuehnii*, causal agent of orange rust of sugarcane, in the United States and western hemisphere. *Plant Dis.* **2008**, *92*, 175–178. [CrossRef] [PubMed]
- Viswanathana, R.; Ashwin, N.M.R. Brown spot of sugarcane: An emerging disease in South Western region in India. *J. Sugarcane Res.* **2020**, *10*, 87–93. [CrossRef]
- Patel, R.R.; Patel, D.D.; Bhatt, J.; Thakor, P.; Triplett, L.R.; Thakkar, V.R. Induction of pre-chorismate, jasmonate and salicylate pathways by *Burkholderia* sp. RR18 in peanut seedlings. *J. Appl. Microbiol.* **2021**, *131*, 1417–1430. [CrossRef] [PubMed]
- Patel, R.R.; Thakkar, V.R.; Subramanian, B.R. A *Pseudomonas guariconensis* strain capable of promoting growth and controlling collar rot disease in *Arachis hypogaea* L. *Plant Soil* **2015**, *390*, 369–381. [CrossRef]
- Wang, X.Y.; Li, J.; Yang, K.; Shan, H.L.; Zhang, R.Y.; Wang, C.M.; Cang, X.Y.; Yin, J.; Luo, Z.M.; Li, W.F.; et al. Evaluation of resistance to brown streak disease in new and main cultivated sugarcane varieties. *Acta Phytopathol. Sin.* **2021**, *51*, 287–293.
- Yen, W.Y.; Chi, C.C. Studies on leaf blight of sugarcane (I). *J. Sugarcane Res. Taiwan* **1952**, *6*, 191–215.
- Hsieh, W.H. The Causal organism of sugarcane leaf blight. *Mycologia* **1997**, *71*, 892–898. [CrossRef]
- Shah, D.A.; Bergstrom, G.C. A rainfall-based model for predicting the regional incidence of wheat seed infection by *Stagonospora nodorum* in New York. *Phytopathology* **2002**, *92*, 511–518. [CrossRef]
- Abeysekara, N.S.; Friesen, T.L.; Keller, B.; Faris, J.D. Identification and characterization of a novel host-toxin interaction in the wheat-*Stagonospora nodorum* pathosystem. *Theor. Appl. Genet.* **2009**, *120*, 117–126. [CrossRef]
- Solomon, P.S.; Lowe, R.G.H.; Tan, K.-C.; Waters, O.D.C.; Oliver, R.P. *Stagonospora nodorum*: Cause of *stagonospora nodorum* blotch of wheat. *Mol. Plant Pathol.* **2006**, *7*, 147–156. [CrossRef]
- Hsieh, W.H.; Comstock, J.C. Leaf Blight. In *A Guide to Sugarcane Diseases*; Rott, P., Bailey, R.A., Comstock, J.C., Croft, A., Saumtally, A.S., Eds.; Production Cirad Publications: Paris, France, 2000; pp. 111–113, ISSN 1251-7224.
- O'Neill, N.R.; Farr, D.F. Miscanthus blight, a new foliar disease of ornamental grasses and sugarcane incited by *Leptosphaeria* sp. and its anamorphic state *Stagonospora* sp. *Plant Dis.* **1996**, *80*, 980–987. [CrossRef]
- Ren, H. Isolation, Genome Assembly and Molecular Detection Method of the Pathogen Causing Sugarcane Leaf Blight. Master's Thesis, Fujian Agriculture and Forestry University, Fuzhou, China, April 2022.

25. Wang, Z.; Ren, H.; Pang, C.; Lu, G.; Xu, F.; Cheng, W.; Que, Y.; Xu, L. An autoploid-suitable polyBSA-seq strategy for screening candidate genetic markers linked to leaf blight resistance in sugarcane. *Theor. Appl. Genet.* **2022**, *135*, 623–636. [[CrossRef](#)] [[PubMed](#)]
26. Wang, Z.; Lu, G.; Wu, Q.; Li, A.; Que, Y.; Xu, L. Isolating QTL controlling sugarcane leaf blight resistance using a two-way pseudo-testcross strategy. *Crop J.* **2022**, *10*, 1131–1140. [[CrossRef](#)]
27. Tedersoo, L.; Albertsen, M.; Anslan, S.; Callahan, B. Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Appl. Environ. Microbiol.* **2021**, *87*, e0062621. [[CrossRef](#)] [[PubMed](#)]
28. Aragona, M.; Haegi, A.; Valente, M.T.; Riccioni, L.; Orzali, L.; Vitale, S.; Luongo, L.; Infantino, A. New-generation sequencing technology in diagnosis of fungal plant pathogens: A dream comes true? *J. Fungi* **2022**, *8*, 737. [[CrossRef](#)]
29. Guiglielmoni, N.; Houtain, A.; Derzelle, A.; Van Doninck, K.; Flot, J.F. Overcoming uncollapsed haplotypes in long-read assemblies of non-model organisms. *BMC Bioinform.* **2021**, *22*, 303. [[CrossRef](#)]
30. Bao, J.; Chen, M.; Zhong, Z.; Tang, W.; Lin, L.; Zhang, X.; Jiang, H.; Zhang, D.; Miao, C.; Tang, H.; et al. PacBio sequencing reveals transposable elements as a key contributor to genomic plasticity and virulence variation in *Magnaporthe oryzae*. *Mol. Plant* **2017**, *10*, 1465–1468. [[CrossRef](#)]
31. Zhong, Z.; Chen, M.; Lin, L.; Han, Y.; Bao, J.; Tang, W.; Lin, L.; Lin, Y.; Somai, R.; Lu, L.; et al. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. *ISME J.* **2018**, *12*, 1867–1878. [[CrossRef](#)]
32. Kelly, A.C.; Ward, T.J. Population genomics of *Fusarium graminearum* reveals signatures of divergent evolution within a major cereal pathogen. *PLoS ONE* **2018**, *13*, e0194616.
33. Alouane, T.; Rimbart, H.; Bormann, J.; Gonzalez-Montiel, G.A.; Loesgen, S.; Schafer, W.; Freitag, M.; Langin, T.; Bonhomme, L. Comparative genomics of eight *Fusarium graminearum* strains with contrasting aggressiveness reveals an expanded open pangenome and extended effector content signatures. *Int. J. Mol. Sci.* **2021**, *22*, 6257. [[CrossRef](#)]
34. Feng, Z.; Hsiang, T.; Liang, X.; Zhang, R.; Sun, G. Draft genome sequence of cumin blight pathogen *Alternaria burnsii*. *Plant Dis.* **2021**, *105*, 1165–1167. [[CrossRef](#)] [[PubMed](#)]
35. Haridas, S.; Albert, R.; Binder, M.; Bloem, J.; LaButti, K.; Salamov, A.; Andreopoulos, B.; Baker, S.E.; Barry, K.; Bills, G.; et al. 101 Dothideomycetes genomes: A test case for predicting lifestyles and emergence of pathogens. *Stud. Mycol.* **2020**, *96*, 141–153. [[CrossRef](#)] [[PubMed](#)]
36. Zeiner, C.A.; Purvine, S.O.; Zink, E.M.; Pasa-Tolic, L.; Chaput, D.L.; Haridas, S.; Wu, S.; LaButti, K.; Grigoriev, I.V.; Henrissat, B.; et al. Comparative analysis of secretome profiles of manganese (II)-oxidizing ascomycete fungi. *PLoS ONE* **2016**, *11*, e0157844. [[CrossRef](#)]
37. Hane, J.K.; Lowe, R.G.; Solomon, P.S.; Tan, K.C.; Schoch, C.L.; Spatafora, J.W.; Crous, P.W.; Kodira, C.; Birren, B.W.; Galagan, J.E. Dothideomycete plant interactions illuminated by genome sequencing and EST analysis of the wheat pathogen *Stagonospora nodorum*. *Plant Cell* **2007**, *19*, 3347–3368. [[CrossRef](#)]
38. Ranallo-Benavidez, T.R.; Jaron, K.S.; Schatz, M.C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **2020**, *11*, 1432. [[CrossRef](#)] [[PubMed](#)]
39. Kokot, M.; Dlugosz, M.; Deorowicz, S. KMC 3: Counting and manipulating k-mer statistics. *Bioinformatics* **2017**, *33*, 2759–2761. [[CrossRef](#)] [[PubMed](#)]
40. Hu, J.; Fan, J.; Sun, Z.; Liu, S. NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **2020**, *36*, 2253–2255. [[CrossRef](#)]
41. Manni, M.; Berkeley, M.R.; Seppey, M.; Simao, F.A.; Zdobnov, E.M. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **2021**, *38*, 4647–4654. [[CrossRef](#)]
42. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)]
43. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
44. Jung, Y.; Han, D. BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics* **2022**, *38*, 2404–2413. [[CrossRef](#)] [[PubMed](#)]
45. Nurk, S.; Koren, S.; Rhie, A.; Rautiainen, M.; Bizkadze, A.V.; Mikheenko, A.; Vollger, M.R.; Altemose, N.; Uralsky, L.; Gershman, A.; et al. The complete sequence of a human genome. *Science* **2022**, *376*, 44–53. [[CrossRef](#)] [[PubMed](#)]
46. Flynn, J.M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A.G.; Feschotte, C.; Smit, A.F. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9451–9457. [[CrossRef](#)]
47. Bruna, T.; Hoff, K.J.; Lomsadze, A.; Stanke, M.; Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* **2021**, *3*, lqaa108. [[PubMed](#)]
48. Hoff, K.J.; Stanke, M. Predicting genes in single genomes with AUGUSTUS. *Curr. Protoc. Bioinform.* **2019**, *65*, e57. [[CrossRef](#)]
49. Bruna, T.; Lomsadze, A.; Borodovsky, M. GeneMark-EP+: Eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom. Bioinform.* **2020**, *2*, lqaa026. [[CrossRef](#)]
50. Chan, P.P.; Lin, B.Y.; Mak, A.J.; Lowe, T.M. tRNAscan-SE 2.0: Improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **2021**, *49*, 9077–9096. [[CrossRef](#)]
51. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013**, *29*, 2933–2935. [[CrossRef](#)]

52. Kalvari, I.; Nawrocki, E.P.; Ontiveros-Palacios, N.; Argasinska, J.; Lamkiewicz, K.; Marz, M.; Griffiths-Jones, S.; Toffano-Nioche, C.; Gautheret, D.; Weinberg, Z.; et al. Rfam 14: Expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.* **2021**, *49*, D192–D200. [[CrossRef](#)]
53. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [[CrossRef](#)]
54. Cantalapiedra, C.P.; Hernandez-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [[CrossRef](#)] [[PubMed](#)]
55. Aramaki, T.; Blanc-Mathieu, R.; Endo, H.; Ohkubo, K.; Kanehisa, M.; Goto, S.; Ogata, H. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **2020**, *36*, 2251–2252. [[CrossRef](#)] [[PubMed](#)]
56. Buchfink, B.; Reuter, K.; Drost, H.G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **2021**, *18*, 366–368. [[CrossRef](#)] [[PubMed](#)]
57. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **2013**, *41*, e121. [[CrossRef](#)]
58. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sonderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [[CrossRef](#)]
59. Moller, S.; Croning, M.D.; Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **2001**, *17*, 646–653. [[CrossRef](#)]
60. Sperschneider, J.; Dodds, P.N. EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and Oomycetes. *Mol. Plant Microbe Interact.* **2022**, *35*, 146–156. [[CrossRef](#)]
61. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. [[CrossRef](#)]
62. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [[CrossRef](#)]
63. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
64. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)] [[PubMed](#)]
65. Letunic, I.; Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [[CrossRef](#)] [[PubMed](#)]
66. Wang, Y.; Kang, H.; Yao, J.; Li, Z.; Xia, X.; Zhou, S. An improved genome sequence resource of *Bipolaris maydis*, causal agent of Southern corn leaf blight. *Phytopathology* **2022**, *112*, 1386–1390. [[CrossRef](#)]
67. O’Connell, R.J.; Thon, M.R.; Hacquard, S.; Amyotte, S.G.; Kleemann, J.; Torres, M.F.; Damm, U.; Buiate, E.A.; Epstein, L.; Alkan, N.; et al. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat. Genet.* **2012**, *44*, 1060–1065. [[CrossRef](#)]
68. Syauqi, J.; Chen, R.K.; Cheng, A.H.; Wu, Y.F.; Chung, C.L.; Lin, C.C.; Chou, H.P.; Wu, H.Y.; Jian, J.Y.; Liao, C.T.; et al. Surveillance of rice blast resistance effectiveness and emerging virulent isolates in Taiwan. *Plant Dis.* **2022**. [[CrossRef](#)] [[PubMed](#)]
69. Reddy, B.; Mehta, S.; Prakash, G.; Sheoran, N.; Kumar, A. Structured framework and genome analysis of *Magnaporthe grisea* inciting pearl millet blast disease reveals versatile metabolic pathways, protein families, and virulence factors. *J. Fungi* **2022**, *8*, 614. [[CrossRef](#)] [[PubMed](#)]
70. Ellwood, S.R.; Liu, Z.; Syme, R.A.; Lai, Z.; Hane, J.K.; Keiper, F.; Moffat, C.S.; Oliver, R.P.; Friesen, T.L. A first genome assembly of the barley fungal pathogen *Pyrenophora Teres F. Teres*. *Genome Biol.* **2010**, *11*, R109. [[CrossRef](#)]
71. Grandaubert, J.; Lowe, R.G.T.; Soyer, J.L.; Schoch, C.L.; Van de Wouw, A.P.; Fudal, I.; Robbertse, B.; Lapalu, N.; Links, M.G.; Ollivier, B.; et al. Transposable element-assisted evolution and adaptation to host plant within the *Leptosphaeria maculans*-*Leptosphaeria biglobosa* species complex of fungal pathogens. *BMC Genom.* **2014**, *15*, 891. [[CrossRef](#)]
72. Devanna, B.N.; Jain, P.; Solanke, A.U.; Das, A.; Thakur, S.; Singh, P.K.; Kumari, M.; Dubey, H.; Jaswal, R.; Pawar, D.; et al. Understanding the dynamics of blast resistance in rice-*Magnaporthe oryzae* interactions. *J. Fungi* **2022**, *8*, 584. [[CrossRef](#)]
73. Hane, J.K.; Williams, A.; Oliver, R.P. Genomic and comparative analysis of the class Dothideomycetes. In *The Mycota*; Poggeler, S., Wostemeyer, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 14, pp. 205–226.
74. Bertazzoni, S.; Jones, D.A.B.; Phan, H.T.; Tan, K.C.; Hane, J.K. Chromosome-level genome assembly and manually-curated proteome of model necrotroph *Parastagonospora nodorum* Sn15 reveals a genome-wide trove of candidate effector homologs, and redundancy of virulence-related functions within an accessory chromosome. *BMC Genom.* **2021**, *22*, 382. [[CrossRef](#)]
75. Faino, L.; Seidl, M.F.; Shi-Kunne, X.; Pauper, M.; van den Berg, G.C.; Wittenberg, A.H.; Thomma, B.P. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* **2016**, *26*, 1091–1100. [[CrossRef](#)] [[PubMed](#)]
76. Huang, M.; Ma, Z.; Zhou, X. Comparative genomic data provide new insight on the evolution of pathogenicity in *Sporothrix* species. *Front. Microbiol.* **2020**, *11*, 565439. [[CrossRef](#)] [[PubMed](#)]



77. Jayasuriya, H.; Silverman, K.C.; Zink, D.L.; Jenkins, R.G.; Sanchez, M.; Pelaez, F.; Vilella, D.; Lingham, R.B.; Singh, S.B. Clavarinic acid: A triterpenoid inhibitor of farnesyl-protein transferase from *Clavariadelphus truncatus*. *J. Nat. Prod.* **1998**, *61*, 1568–1570. [[CrossRef](#)] [[PubMed](#)]
78. Lingham, R.B.; Silverman, K.C.; Jayasuriya, H.; Kim, B.M.; Amo, S.E.; Wilson, F.R.; Rew, D.J.; Schaber, M.D.; Bergstrom, J.D.; Koblan, K.S.; et al. Clavarinic acid and steroidal analogues as Ras-and FPP-directed inhibitors of human farnesyl-protein transferase. *J. Med. Chem.* **1998**, *41*, 4492–4501. [[CrossRef](#)]
79. Godio, R.P.; Fouces, R.; Martin, J.F. A squalene epoxidase is involved in biosynthesis of both the antitumor compound clavarinic acid and sterols in the basidiomycete *H. sublateritium*. *Chem. Biol.* **2007**, *14*, 1334–1346. [[CrossRef](#)]
80. Du, X.; Li, H.; Qi, J.; Chen, C.; Lu, Y.; Wang, Y. Genome mining of secondary metabolites from a marine-derived *Aspergillus terreus* B12. *Arch. Microbiol.* **2021**, *203*, 5621–5633. [[CrossRef](#)]
81. Zhu, S.; Yan, Y.; Qu, Y.; Wang, J.; Feng, X.; Liu, X.; Lin, F.; Lu, J. Role refinement of melanin synthesis genes by gene knockout reveals their functional diversity in *Pyricularia oryzae* strains. *Microbiol. Res.* **2021**, *242*, 126620. [[CrossRef](#)]
82. Freitas, D.F.; Rocha, I.M.; Vieira-da-Motta, O.; de Paula Santos, C. The role of melanin in the biology and ecology of nematophagous Fungi. *J. Chem. Ecol.* **2021**, *47*, 597–613. [[CrossRef](#)]
83. Members, C.-N.; Partners. Database resources of the National Genomics Data Center, China National Center for Bioinformatics in 2021. *Nucleic Acids Res.* **2021**, *49*, D18–D28.