Article

Genome Analysis of *Phytophthora nicotianae* JM01 Provides Insights into Its Pathogenicity Mechanisms

Xiao-Long Yuan [1,2,†] , Cheng-Sheng Zhang [1,2,*,†] , Fan-Yu Kong [1,2], Zhong-Feng Zhang [1,2] and Feng-Long Wang [1,2,*]

1 Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao 266101, China; yuanxiaolong@caas.cn (X.-L.Y.); kongfanyu@caas.cn (F.-Y.K.); zhangzhongfeng@caas.cn (Z.-F.Z.)
2 Special Crops Research Center of Chinese Academy of Agricultural Sciences, Qingdao 266101, China
* Correspondence: zhangchegnsheng@caas.cn (C.-S.Z.); wangfenglong@caas.cn (F.-L.W.); Tel.: +86-532-88701035 (C.-S.Z. & F.-L.W.)
† These authors contributed equally to this work.

Abstract: *Phytophthora nicotianae* is a widely distributed plant pathogen that can cause serious disease and cause significant economic losses to various crops, including tomatoes, tobacco, onions, and strawberries. To understand its pathogenic mechanisms and explore strategies for controlling diseases caused by this pathogen, we sequenced and analyzed the whole genome of *Ph. nicotianae* JM01. The *Ph. nicotianae* JM01 genome was assembled using a combination of approaches including shotgun sequencing, single-molecule sequencing, and the Hi-C technique. The assembled *Ph. nicotianae* JM01 genome is about 95.32 Mb, with contig and scaffold N50 54.23 kb and 113.15 kb, respectively. The average GC content of the whole-genome is about 49.02%, encoding 23,275 genes. In addition, we identified 19.15% of interspersed elements and 0.95% of tandem elements in the whole genome. A genome-wide phylogenetic tree indicated that *Phytophthora* diverged from *Pythium* approximately 156.32 Ma. Meanwhile, we found that 252 and 285 gene families showed expansion and contraction in *Phytophthora* when compared to gene families in *Pythium*. To determine the pathogenic mechanisms *Ph. nicotianae* JM01, we analyzed a suite of proteins involved in plant–pathogen interactions. The results revealed that gene duplication contributed to the expansion of Cell Wall Degrading Enzymes (CWDEs) such as glycoside hydrolases, and effectors such as Arg-Xaa-Leu-Arg (RXLR) effectors. In addition, transient expression was performed on *Nicotiana benthamiana* by infiltrating with *Agrobacterium tumefaciens* cells containing a cysteine-rich (SCR) protein. The results indicated that SCR can cause symptoms of hypersensitive response. Moreover, we also conducted comparative genome analysis among four *Ph. nicotianae* genomes. The completion of the *Ph. nicotianae* JM01 genome can not only help us understand its genomic characteristics, but also help us discover genes involved in infection and then help us understand its pathogenic mechanisms.

Keywords: *Phytophthora nicotianae*; genome; divergence time; pathogenic mechanisms; comparative genomics

1. Introduction

Oomycetes are a kind of ubiquitous filamentous microorganisms that can cause various destructive diseases and cause economic losses of agricultural crops, ornamental plants, and natural plant populations [1–3]. Horizontal and endosymbiotic gene transfer events, gene fusions and duplications shaped and diversified the genomes of species in oomycetes [4,5]. Among which, species that belong to the genera *Phytophthora*, *Pythium*, and *Albugo* are the most diverse and destructive pathogens [1]. Therefore, many studies have been conducted to understand their virulence mechanisms, diversities and evolutionary positions, to identify targets for drugs, and crop protection chemicals to control disease. The virulence mechanisms and evolution of oomycetes have mainly been examined through transcripts and whole-genome sequences [6–9]. For example, the divergence

times of oomycetes have been estimated under distinct molecular clock models, indicating the oomycete diverged about 430–400 MYA. Additionally, the divergence time of two major lineages Saprolegniales and Peronosporales in oomycetes were placed in the early Mesozoic (~225–190 MYA) [6]. Meanwhile, transcriptomics and comparative transcriptomic analyses have been performed among species in oomycetes to identify their transcriptional remodeling when the host was exposed to infection and chemical environments. For example. RNA-seq analysis of *Phytophthora capsici* identified stage-specific genes including effector families and metabolic pathways, revealing that these proteins are important in the infection process [10]. In addition, transcriptome analysis of three life stages of *Ph. litchii* predicted 490 pathogenicity-related genes [11]. Similar, RNA-seq of *Aphanomyces euteiches* showed that the adaptability to plant hosts was related to the expression of specialized secretions [12]. Importantly, studies of oomycetes have been advanced by the completion of their genome sequences. For example, the genome sequence of *Saprolegnia parasitica* revealed that some pathogenesis-related genes in this species originated from lateral gene transfer (LGT), thereby indicating the evolutionary adaptation to its hosts [13]. The completion of the genome of *Pythium ultimum* helped us understand a new pathogenic mechanism, which characterized by lacking crucial effectors such as Arg-Xaa-Leu-Arg (RXLR) effectors and Crinkler genes and gaining a novel YxSL family [8]. *Py. insidiosum* can cause infectious disease for both humans and animals, and its virulence and pathogenicity were previously examined at the genome level [14]. Moreover, a 38.7 Mb *Py. destruens* genome was generated by whole-genome shotgun sequencing technique [15]. *Phytophthora* is a notorious oomycete genus that contains more than 100 species and can cause damage to a range of economically important agricultural and horticultural plants [2]. Analysis of the genome of *Ph. sojae* revealed genes related to plant infections such as transporters, protein toxins and proteinase inhibitors [16]. *Ph. infestans* is a model oomycete that causes destructive disease of potatoes, resulting in huge economic losses. In 2009, the *Ph. infestans* genome revealed the notable rapid expansion and evolution of effector genes [6]. The 64 Mb genome of the pepper and cucurbits pathogen *Ph. capsica* was published in 2012 [17], while its genome size was corrected to 100.5 Mb by the long-read sequencing method in 2021 [18]. Additionally, the genomes of cacao black pod rot pathogens *Ph. megakarya* and *Ph. palmivora* were also assembled in 2017 [19]. In 2018, Vetukuri et al. reported the genome sequences of *Ph. plurivora* and *Ph. colocasiae*, and analyzed their pathogenicity-related genes, such as RXLR, Crinkle, and hypersensitive response effectors [20]. In addition, Yang et al. generated a high-quality genome assembly of *Ph. vignae* strain PSY2020 using Oxford Nanopore Technologies (ONT) sequencing and identified 924 candidate RXLR effectors [21]. In addition, comparative genomic analysis among species in oomycetes showed that genes related to secreted effectors and CWDZs exhibited expansion in *Phytophthora*, while genes encoding proteolytic degradation and signal transduction were increased in *Pythium*, which provided important insights into the evolutionary history of oomycetes and their mechanisms of pathogenesis [22].

Phytophthora nicotianae is widely distributed in soil and irrigation water, and is one of the most important pathogens in the Peronosporales, which can survive in field soil for several years [23]. For tobacco, it can cause black shank disease in different tissues, and cause root rot, leaf wilting, black stems and even death, which greatly reduces tobacco production and causes significant economic losses [24]. Records of the private company indicated that the occurs of *Ph. nicotianae* and can cause losses of about 65% or more in Colombia. Additionally, the average annual economic loss caused by *Ph. nicotianae* in China is more than one hundred million yuan [25]. Previous studies have identified four physiological races of this pathogen (0, 1, 2 and 3) in major tobacco cultivation areas worldwide [26]. Among these races, the most common strain, race 0 has been the predominant race in the flue-cured area of North Carolina since 1931 [27]. Race 1 is pathogenic on *N. plumbaginifolia*, but its fitness was not equivalent to race 0, and it can be induced under intense selection pressure of continuous breeding of race 0 resistant varieties [28]. Races 2 and 3 have been reported in South Africa and Connecticut, respectively [29,30]. Previous studies of *Ph.*

nicotianae mainly focused on morphological and physiological identification, pathology, evolution, and prevention measures [31–33]. However, this disease is difficult to control via chemical means. Thus, the breeding of pathogen-resistant cultivated lines currently presented is the major control strategy to control this pathogen. Genome information can yield tremendous information that helps us understand its evolutionary position and pathogenic mechanisms. Currently, there are already three assembled genomes of *Ph. nicotianae* in NCBI database (*Ph. nicotianae* race 0, race 1 and BL) [26]. The genomes *Ph. nicotianae* race 0 and race 1 were published in 2016 by a combination of PacBio single-molecular real-time sequencing and second-generation sequencing methods, with a genome size of 81 Mb and 71 Mb, respectively. Genome analysis showed that the *Ph. nicotianae* race 0 and race 1 contain expanded ATP-binding cassette transporter gene family when compared with those in *Ph. infestans*. This result suggests that the ATP-binding cassette transporters plays important roles in the adaptive evolution of *Ph. nicotianae* to the host [26]. In addition, Liu et al. (2016). found that the difference in RXLR effector gene numbers between race 0 and race 1 may be related with their pathogenicities in cultivated tobacco [26]. The genome size of *Ph. nicotianae* BL was about 106.75 Mb. However, no systematic studies have been performed for their divergence time and virulence mechanisms.

Therefore, we isolated *Ph. nicotianae* JM01 from tobacco and sequenced its whole genome and analyzed its secreted proteins and effectors to investigate its pathogenic mechanisms. We estimated the divergence time of *Ph. nicotianae* JM01 based on molecular time markers and compared its genome with some closely related oomycetes. These results help us understand the pathogenic mechanism of *Ph. nicotianae* JM01 and provide a theoretical basis for future disease prevention and control.

2. Results

2.1. Genome Sequencing, Assembly and Validation

Sequencing reads that were generated from Illumina and PacBio platforms were used for assembling the *Ph. nicotianae* JM01 genome after quality control. A total of ~22.14 Gb and ~5.32 Gb of clean data were obtained from shotgun sequencing platform (21.20 Gb for 170 bp, 10.68 Gb for 500 bp, 3.50 Gb for 2000 bp and 1.42 Gb for 6000 bp libraries) and single-molecule sequencing platform, respectively. For the Hi-C library, we produced 28.37 Gb clean data. First, we obtained a draft genome of *Ph. nicotianae* JM01, with a total length of 59.05 Mb, with an N50 contig and scaffold length of 54.01 kb and 100.23 kb, respectively. After assembling the contigs, we used mate-pair sequencing data and PacBio sequences to generate scaffolds. Combined with the Hi-C data, we obtained a genome of approximately 95.68 Mb in size, consisting of 1446 scaffolds. The contig and scaffold N50 are 54.23 kb and 176.88 kb, respectively (Table 1). To assess the quality of the assembled genome of *Ph. nicotianae* JM01, we analyzed the completeness of the gene sets using BUSCO. The results showed that 92.37% of the eukaryotic single-copy genes were detected in the genome of *Ph. nicotianae* JM01 (Supplementary Figure S1).

Table 1. Genomic features of *Phytophthora nicotianae*.

Categories	Scaffold	Contig
Total Number	2105	3868
Total length (bp)	95,318,700	58,288,754
Average Length (bp)	29,463.56	15,069.48
N50 Length (kb)	113.15	54.23
N90 Length (kb)	28.94	8.54
Maximum Length (bp)	540,321	383,516
Minimum Length (bp)	1002	200
GC content	49.02%	49.78%

2.2. Repeat Elements

Repeat elements constituted 20.10% of the *Ph. nicotianae* JM01 genome, with 19.15% interspersed repeats and 0.95% tandem repeat sequences, respectively. Among the identified interspersed repeats, we found that 9461 long terminal repeated LTRs (3,660,470 bp) covered 7.62% of the whole genome, among which 7785 (6.38%) belonged to the Gypsy superfamily and 1179 (0.83%) belonged to the Copia superfamily. Additionally, we identified 214 (0.16%) LTRs that belonged to the Ngaro superfamily. DNA elements, including MuLE-MuDR, TcMar-Tc1, and PiggyBac, comprised 3.52% of the genome (Table 2).

Table 2. Repeat elements in *Ph. nicotianae* genome.

Class	Type	Number of Elements	Length Occupied	Percentage of Sequence
		DNA elements		
	Crypton	535	320,797	0.34%
	Maverick	110	143,341	0.15%
	MuLE-MuDR	1243	587,998	0.62%
	MuLE-NOF	535	60,260	0.06%
	PIF-Harbinger	294	231,089	0.24%
	PiggyBac	450	257,994	0.27%
	Sola	275	91,598	0.10%
	TcMar-ISRm11	791	398,902	0.42%
	Merlin	198	88,846	0.09%
	TcMar-Pogo	201	54,855	0.06%
	TcMar-Ant1	272	138,354	0.15%
Interspersed	TcMar-Tc1	807	449,528	0.47%
repeats	TcMar-Tc2	469	215,152	0.23%
	TcMar-Stowawa	197	113,580	0.12%
	hAT-Ac	324	136,041	0.14%
	hAT-Tag1	176	68,793	0.07%
	DNA	543	303,342	0.32%
		LTR [#]		
	LTR	283	238,869	0.25%
	Copia	1179	787,427	0.83%
	Gypsy	7785	6,079,720	6.38%
	Ngaro	214	148,614	0.16%
		RC [$]		
	Helitron	654	444,976	0.47%
	Unknown	20279	6,888,937	7.23%
	Microsatellite	717	52,963	0.06%
Tandem repeats	Minisatellite	7437	598,915	0.63%
	Satellite	576	255,356	0.27%
Total		46544	19,156,247	20.10%

[#] long terminal repeats; [$] rolling circle.

After repeat masking, de novo prediction, homology searches, and RNA-based methods were used to predict gene structures of *Ph. nicotianae* JM01. *De novo* prediction using AUGUSTUS 3.0.2 predicted 17,325 gene models for *Ph. nicotianae* JM01 [34]. For homologous annotation, *Ph. nicotianae* JM01 genome scaffolds were used as queries to search for a database containing protein sequences from seven *Phytophthora* species (*Ph. capsici*, *Ph. ramorum*, *Ph. infestans*, *Ph. parasitica*, *Ph. cinnamomic*, *Ph. nicotianae* race 0 and *Ph. sojae*) and accordingly identified 45,276 homologous protein sequences. We also mapped the cDNA and ESTs of *Ph. nicotianae* to the genome using BLAST and then assembled using PASA [6]. We totally predicted 15,652 gene models. Finally, EVM was used to integrate the above results and obtain a gene dataset comprising 23,275 protein-encoding genes. Most of these genes contained few introns, with an average of 1.59 introns per gene, and were relatively short (~157.36 bp). The coding exons and coding sequences of the *Ph. nicotianae* JM01 genome showed average lengths of 592.77 bp and 1789.27 bp, respectively.

To further determine the functions of the protein-coding genes in *Ph. nicotianae* JM01, these protein sequences were aligned against several public databases. We accordingly identified 23,130 genes that had at least one hit in the NR database, which have high homology with species in the genus *Phytophthora* (99.38%). The top five species for which we obtained the best BLAST hits for *Ph. nicotianae* JM01 are *Ph. parasitica*, *Ph. infestans*, *Ph. sojae*, *Ph. halstedii,* and *Albugo laibachii* in NR database (Supplementary Figure S2). Based on the results of GO analyses, we successfully assigned a total of 6592 protein-coding genes to GO glossaries with 23,551 functional terms (Supplementary Figure S3). KEGG analysis revealed that 2367 genes have KO annotations (Supplementary Table S2), and a total of 337 KEGG metabolic pathways were annotated in the *Ph. nicotianae* JM01 genome (Supplementary Figure S4). Furthermore, we found that 6550 genes in the *Ph. nicotianae* JM01 genome could be assigned to 1704 CDD superfamilies. Moreover, Pfam annotation showed that a total of 9789 genes in the *Ph. nicotianae* JM01 genome were associated with 30,927 domains.

2.3. Gene Family Expansion and Contraction

To compare gene families and contents among the different species, we performed orthology analysis. The results indicated the expansion and contraction of 574 and 125 gene families, respectively, in the *Ph. nicotianae* JM01 genome when compared with that of *Hyaloperonospora arabidopsidis*. In addition, we identified 252 and 285 gene families in *Phytophthora* that showed expansion and contraction when compared with those in *Pythium*. In *Phytophthora*, we found that among the gene families identified, 136 (containing 728 genes), 13 (127 genes), 24 (216 genes), 15 (17 genes), 48 (1959 genes), 39 (812 genes), 14 (14 genes), 36 (2514 genes), 31 (222 genes) and 32 (1595 genes) were uniquely expanded in *Ph. cinnamomi*, *Ph. capsici*, *Ph. nicotianae* JM01, *Ph. nicotianae* race 1, *Ph. nicotianae* BL, *Ph. infestans*, *Ph. nicotianae* race 0, *Ph. parasitica*, *Ph. ramorum,* and *Ph. sojae*, respectively. In *Ph. nicotianae* JM01, we identified 417 and 994 gene families that were expanded and contracted, respectively.

2.4. Phylogenetic Relationship and Divergence Time Analyses

To elucidate the evolutionary position of *Ph. nicotianae* JM01, we performed phylogenetic relationships and divergence time analysis based on 46 single-copy genes. The constructed phylogenetic tree showed that *T. thermophila*, used as an outgroup, was located at the basal position of the whole tree (Figure 1). We identified four distinct lineages in the tree: the brown algae, diatoms, Saprolegniales and Peronosporales. A clade united brown algae and diatoms together, which emerged as a sister group to the clade containing Saprolegniales and Peronosporales. Among the Peronosporales, we found that *Ph. nicotianae* initially clustered with *Ph. parasitica*. A further clade placed the genera *Phytophthora*, *Hyaloperonospora*, *Pythium*, and *Albugo* together. In addition, we found that *Pythium vexans* was located close to a clade containing *Phytophthora* and *Hyaloperonospora*, and then clustered with the genus *Pythium*. Furthermore, our relaxed molecular clock analysis indicated the *T. thermophila* initially appeared at 705.53 Ma (95% highest posterior density (HPD) = 637.62–789.30 Ma). Similarly, we determined that brown algae, diatoms, and Saprolegniales initially appeared at 532.18, 220.91, and 102.12 Ma (95% HPD = 466.62–572.21, 197.03–245.15 and 91.93–112.61 Ma), respectively. We also estimated that the ancestor of Peronosporales originated in the Carboniferous, approximately 348.26 (327.73–385.68) Ma. Among the genera in the Peronosporales, the divergence events were involved in the emergence of *Hyaloperonospora* (87.00(78.66–92.77) Ma), *Pythium* (230.38 (212.34–255.02) Ma) and *Albugo* (293.13 (269.48–318.74) Ma), and we found that species of *Phytophthora* shared a common ancestor at approximately ~58.21 (53.91–63.12) Ma. Additionally, the results showed that all the four strains of *Ph. nicotianae* diverged from each other approximately at ~2.96 (2.38–3.65) Ma.
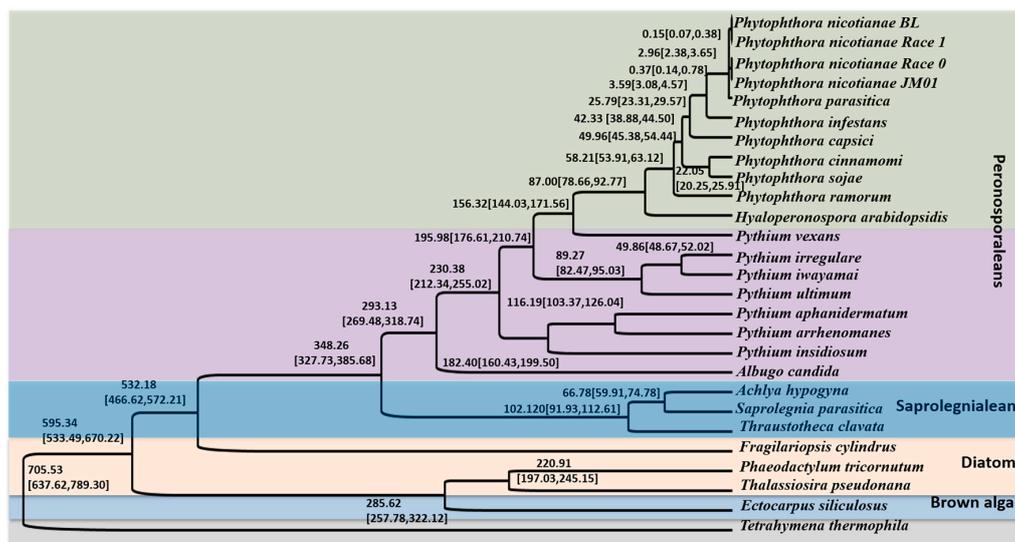
Figure 1. Phylogenetic relationships of *Ph. nicotianae* JM01and other species based on the dataset of common single copy genes. This phylogenetic tree was generated using Bayesian inference in BEAST v1.7.5. The time scale was set according to the divergence time between *T. clavata* and *Py. infestans* (~363 Ma). The divergence time of each species was marked above each branch with 95% confidence level. Different colors represent different groups.

2.5. Analysis of Pathogenicity Mechanisms

The first stage for *Phytophthora* to infect the host is to destruct their cell wall polysaccharide components using carbohydrate-active enzymes (CAZys), which also promotes their utilization of nutrients in the host cell. Our analysis indicated that glycosyl transferases (GTs), glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), auxiliary activities (AA) and the carbohydrate-binding module (CBL) have been expanded in the *Ph. nicotianae* JM01 genome (Figure 2A). Based on the CAZy database, we identified 44 unique carbohydrate enzymes in infected hosts that are unique to the *Ph. nicotianae* JM01 genome. Moreover, AA3, AA4, AA6, AA7, CBM1, CBM13, CBM47, CE8, CH1, CH10, CH109, CH12, CH131, CH140, CH16, CH17, CH28, CH3, CH30, CH43, CH16, CH17, CH28, CH3, CH30, CH43, CH5, CH72, GT1, GT31, PL1 and PL3 underwent significant expansion in *Ph. nicotianae* JM01 genome. Analysis of the phylogenetic relationships of GH genes revealed that the expansion of these genes duplicated in the *Ph. nicotianae* JM01 genome (Figure 2B).

Elicitins can induce a hypersensitive response in some species due to its structures. We observed an expansion of these proteins in *Phytophthora* genomes compared with those in *Pythium* genomes. We identified 42, 60, 48, 32, 51, 54, 57, 54, 54 and 67 elicitins in *Ph. capsici*, *Ph. cinnamomi*, *Ph. infestans*, *Ph. nicotianae* race 0, *Ph. nicotianae* JM01, *Ph. nicotianae* BL, *Ph. nicotianae* race 1, *Ph. ramorum* and *Ph. sojae*, respectively. Exhibiting cellulose-binding, the elicitor of plant defense responses, and lectin-like activities (CBELs), interferons (INFs), oligopeptide elicitors (OPELs) and Peptides (PEPs) are important elicitors in pathogenic organisms. Our results indicated the presence of 16 CBELs, 36 INFs, 16 OPELs, and 21 PEPs in the *Ph. nicotianae* JM01 genome (Supplementary Table S3), which is comparable to the composition in other species.

Many plant pathogens can produce an abundance of secreted proteins which can infect their hosts' immunity. To explore the pathogenic mechanisms of *Ph. nicotianae* JM01, we selected secreted proteins and effectors from known related species as databases. Finally, we identified 2445 predicted secreted proteins in the *Ph. nicotianae* JM01 genome, which is a much larger number than those found in other *Phytophthora* species to date. Among the secreted proteins in the *Ph. nicotianae* JM01 genome, we identified 211 RXLR effectors, which is broadly comparable to the numbers in other *Phytophthora* species, including *Ph. capsica* (357), *Ph. cinnamomic* (173), *Ph. infestans* (323), *Ph. nicotianae* race 0 (133), *Ph. nicotianae*

race 1 (163), *Ph. nicotianae* BL (156), *Ph. parasitica* (230), *Ph. ramorum* (143), and *Ph. sojae* (264) and considerably more than that in the *H. arabidopsidis* genome (29). Similar to other *Phytophthora* species, the RXLR effectors in *Ph. nicotianae* JM01 contain amino-terminal cell entry domains with RXLR and dEER motifs (Figure 3A–C). For investigating the functions of these genes, we performed phylogenetic analyses to establish their relationships. The results showed that gene duplication played an important role in RXLR genes expansion (Supplementary Figure S5).
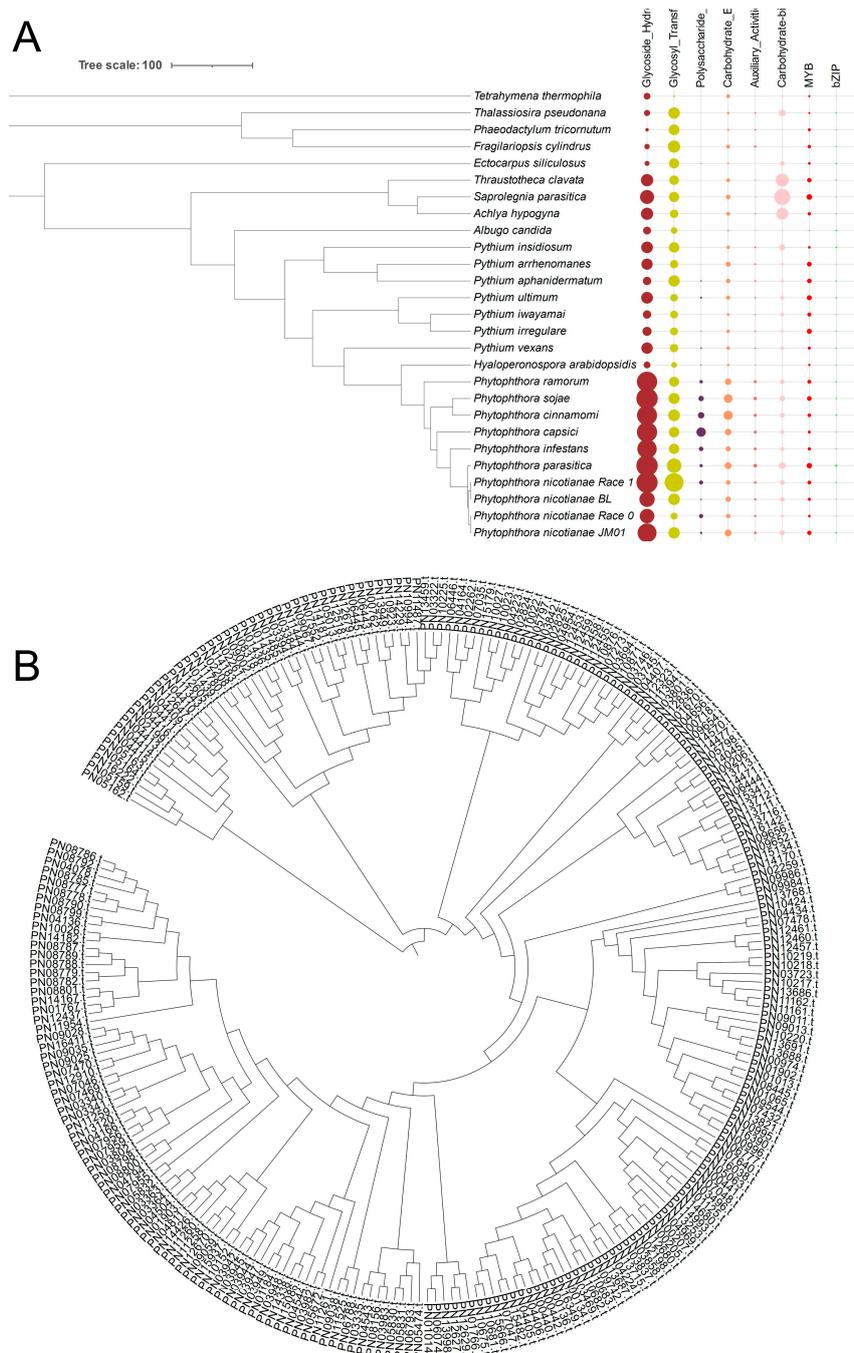

Figure 2. CAZyme annotation and transcription factors of *Ph. nicotianae* JM01 and other closely related genomes. (**A**) The CAZyme annotation and transcription factors results. The tree on the left was generated using Bayesian inference in BEAST v1.7.5. Different circles present glycoside hydrolases, glycosyl transferases, polysaccharide lyases, carbohydrate esterases, auxiliary activities, the carbohydrate-binding module, MYb and bZIP. The sizes of dots on the right present the gene numbers of different CAZyme genes and transcription factors. (**B**) Phylogenetic relationships of glycoside hydrolases (GH) genes in *Ph. nicotianae* JM01.
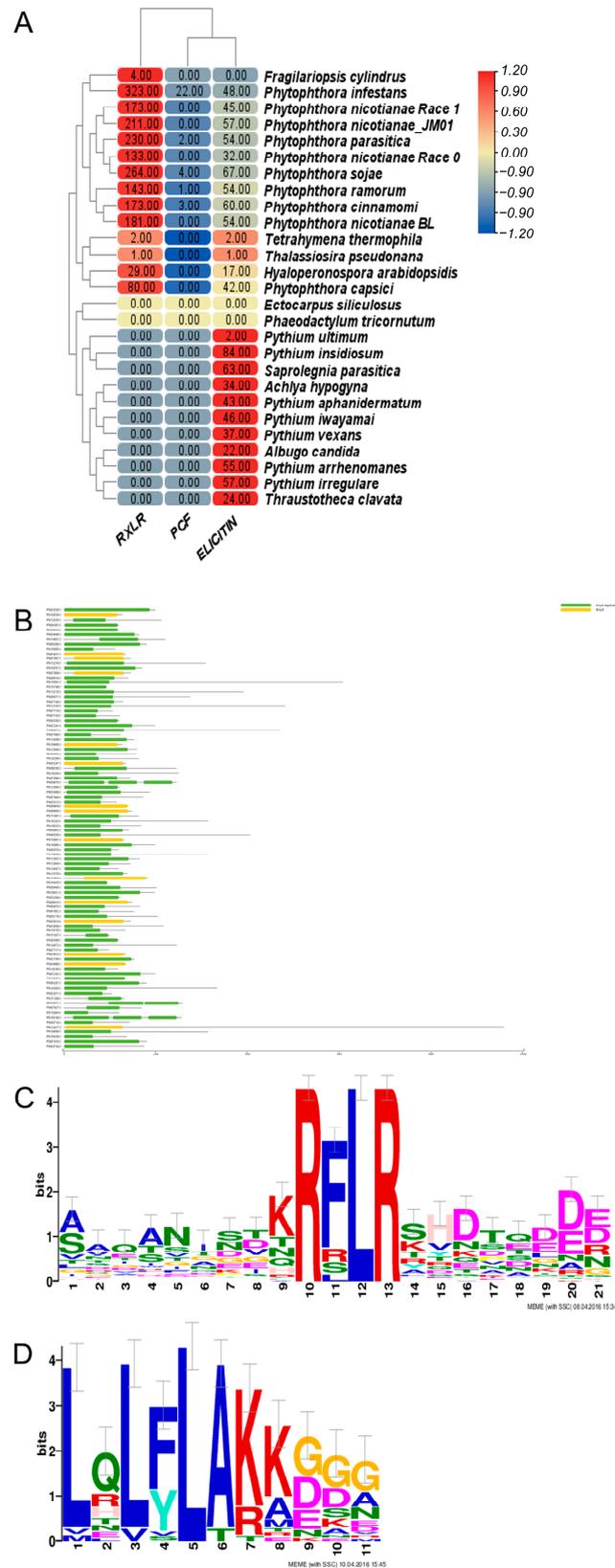
Figure 3. Prediction results of effectors in *Ph. nicotianae* JM01 genome and other species. (**A**) Results of effectors in the *Ph.nicotianae* JM01 genome and other species. A heatmap was used to present the gene numbers of RXLR, PCF and ELICITIN; (**B**) Genes structures of RXLR in *Ph.nicotianae* JM01. Conserved domains were drawn in NCBI; (**C**) Typical architecture of RXLR in *Ph. nicotianae* JM01; (**D**) Typical architecture of CRN in *Ph. nicotianae* JM01.

Crinkler (CRN) represents a type of cytoplasmic effector gene, which encodes a secreted protein containing a conserved amino-terminal LFLAK domain. We identified 24 CRN genes in *Ph. nicotianae* JM01 genome. In contrast to RXLR effectors, the structure of CRN genes showed high conservation in most pathogenic oomycete genomes that have been published to date (Figure 3D). In addition, a total of 97, 37, 196, 22, 24, 26, 28, 86, 79 and 19 CRN genes have been identified in *Ph. cinnamomi*, *Ph. capsici*, *Ph. infestans*, *Ph. nicotianae* race 0, *Ph. nicotianae* JM01, *Ph. nicotianae* BL, *Ph. nicotianae* race 1, *Ph. parasitica*, *Ph. ramorum,* and *Ph. sojae*, respectively. We aligned the amino-terminal regions of all 24 *Ph. nicotianae* JM01 CRN proteins and found that these shared conserved LxLYLAR/K and WL motifs (Supplementary Figure S6). PcF (*Phytophthora cactorum-Fragaria* protein), a cysteine-rich (SCR) protein, was also found only in *Ph. parasitica*, *Ph. infestans*, *Ph. cinnamomi*, *Ph. ramorum* and *Ph. sojae.* In the *Ph. nicotianae* JM01 genome, we identified a SCR protein PnF1, a homolog of PcF. To investigate the function of PnF1, transient expression was performed on *N. benthamiana* by infiltrating with *A. tumefaciens* cells containing PnF1. On day 4 post-treatment, the negative control leaves still remain healthy, whereas those infiltrated with PnF1 showed symptoms of hypersensitive response (Figure 4A). The results showed that the areas were infiltrated with BAX and INF1 showed hypersensitive reaction, while the areas injected with GFP and MgCl$_2$ did not show any symptoms of hypersensitive response (Figure 4B). Cutinase is another cysteine-rich (SCR) protein, which is defined as extracellular serine esterase that can break the ester bonds of cutin. Among *Phytophthora* species, a total of 15 cutinases were found in *Ph. sojae*, only two of which lacked a signal peptide. Interestingly, among the six cutinases in *Ph. capsica,* only one contained a signal peptide. Both *Ph. infestans* and *Ph. ramorum* have four cutinases, all of which contain a signal peptide. Generally, we found that cutinases are highly conserved in the genomes of *Phytophthora* species. In addition to these effectors, we searched for other secreted proteins in the *Ph. nicotianae* JM01 genome, including ATR13, HLE and TOXA. However, we did not find genes coding for these proteins.
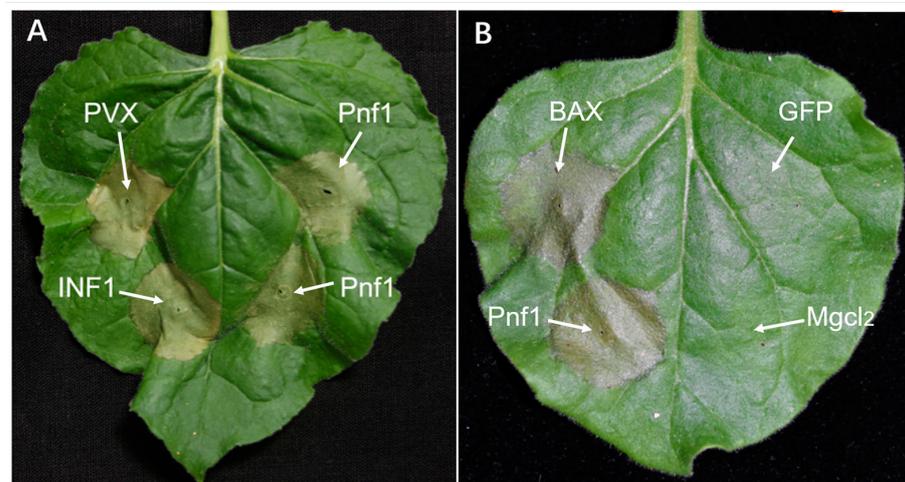

Figure 4. PnF1 from *Ph. nicotianae* JM01 triggers cell death in *N. benthamiana* leaf. (**A**) The recombinant strains of *A. tumefaciens* containing PnF1 were infiltrated into *N. benthamiana* leaf using a needle to detect the function of PnF1. The upper left portion of the leaf in (**A**) is injected with PVX, and the lower left portion of the leaf in (**A**) is injected with INF1. The right portion of the leaf in (**A**) is injected with Pnf1. (**B**) The BCL2-Associated X (BAX) was selected as a positive control, and green fluorescent protein (GFP) and MgCl$_2$ were selected as negative controls. The upper left portion of the leaf in (**B**) is injected with BAX, and the lower left portion of the leaf in (**B**) is injected with Pnf1. The upper right portion of the leaf in (**B**) is injected with GFP and the lower left portion of the leaf in (**B**) is injected with MgCl$_2$.

Transcription factors can bind to DNA and regulate gene expression by either promoting or suppressing transcription. It has been demonstrated that transcription factors play

important roles in pathogenicity by mediating oxidative stress response during plant defense responses and virulence. We identified 19 proteins as potential basic region—leucine zipper (bZIP) proteins (Figure 2A). To further examine the characteristics of these bZIP domains, these sequences were compared with the conventional bZIP domains. As shown in Figure 2A, some of the conserved region comparisons showed differences between the *Ph. nicotianae* JM01 bZIP domains and conventional bZIP domains, especially in DNA recognition regions. The candidate bZIPs were divided into the following six classes: N-R, C-R, R-R, V-R, N-K and others. In addition, we also analyzed the bZIP proteins in other species, and accordingly found that the *Ph. nicotianae* JM01 genome contains a larger number of bZIP proteins than those of the other assessed species. Myb factor is also a type of transcription factor that plays an important role in the process of pathogen infection. In total, we identified 64 Myb in the *Ph. nicotianae* JM01 genome (Figure 2A), much larger than the numbers found in any other *Phytophthora* genome sequenced to date.

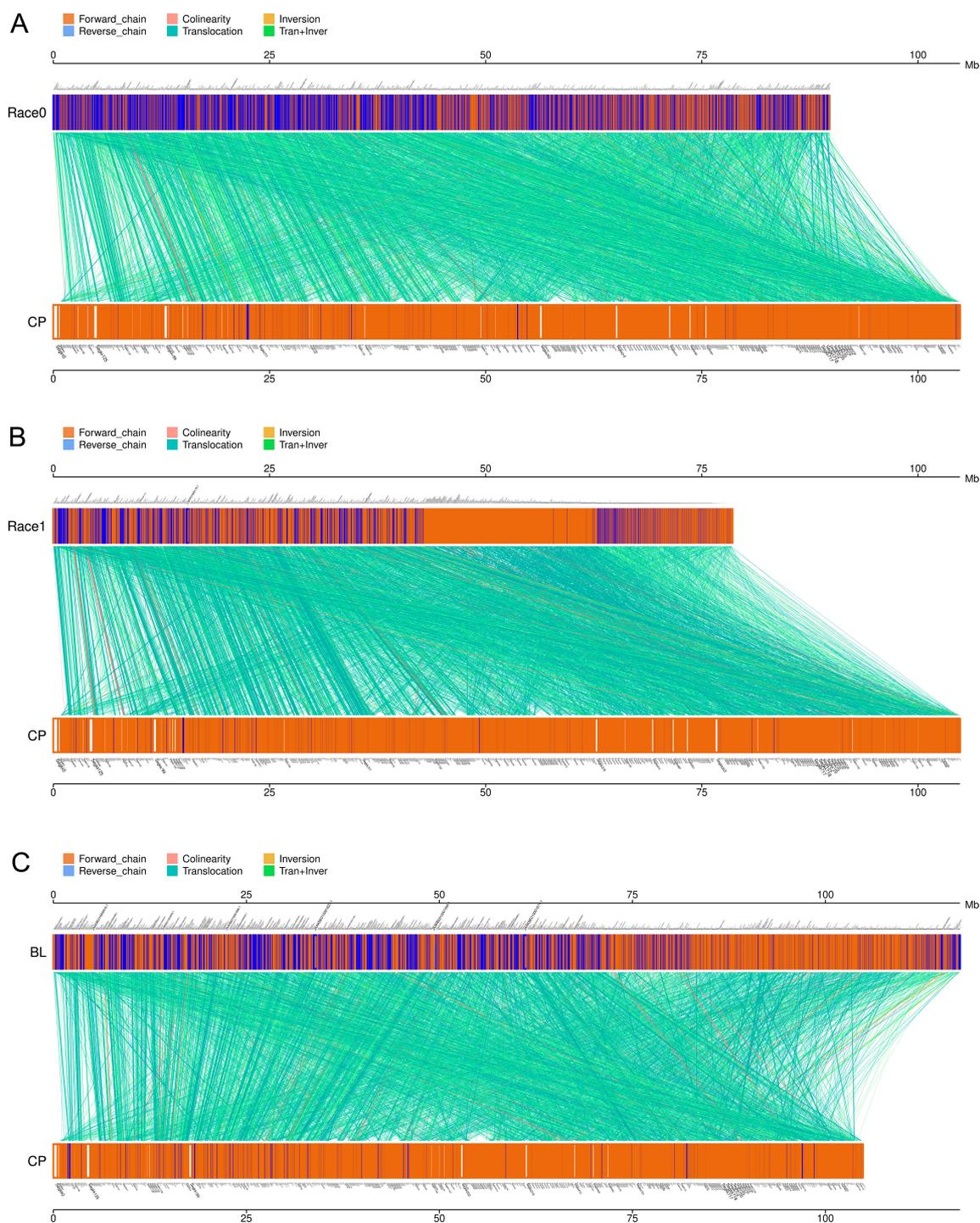### 2.6. Comparative Genomes of Four Ph. nicotianae Genomes

The gene contents of *Ph. nicotianae* JM01 were compared with those of *Ph. nicotianae* race 0, *Ph. nicotianae Ph. nicotianae* race 1, and *Ph. nicotianae* BL. A total of 664 core genes shared by these four strains (Figure 5). These genes were mainly annotated as integrase core domain, growth-arrest specific micro-tubule binding, fibrinogen alpha/beta chain, protein tyrosine kinase, lipopolysaccharide kinase, Viral A-type inclusion protein, etc. Except for the common shared genes, *Ph. nicotianae* JM01, *Ph. nicotianae* race 0, *Ph. nicotianae Ph. nicotianae* race 1, and *Ph. nicotianae* BL contained 680, 2732, 6612 and 991 specific genes, respectively.
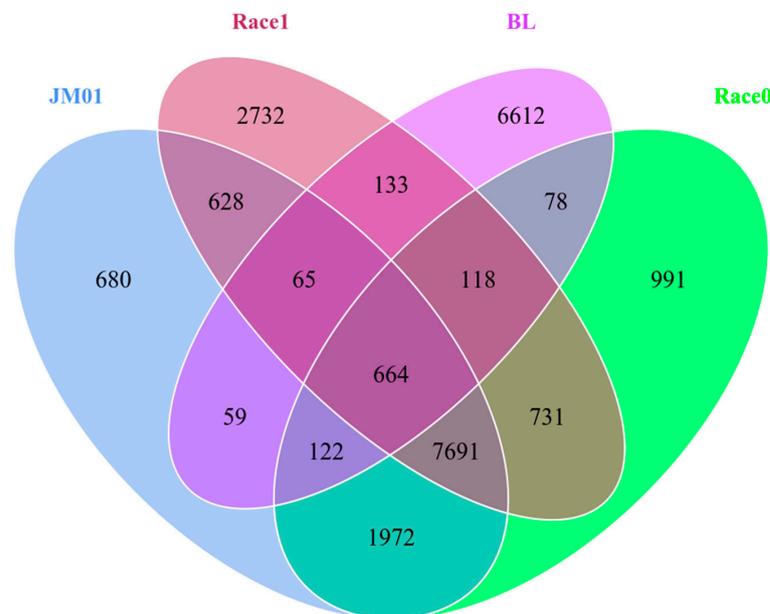
Figure 5. Venn diagram showing the core and pan genes among four *Ph. nicotianae* genomes. Names of each species were listed beside the diagram component. *Ph. nicotianae* JM01 (JM01); *Ph. nicotianae* race 0 (race 0); *Ph. nicotianae* race 1 (race 1); *Ph. nicotianae* BL (BL).

To investigate the genomic structure in four *Ph. nicotianae* genomes, MUMmer software (Version 3.23) was used to analyze the genome synteny. The syntenic regions comparison showed that these genomes shared common syntenic regions. However, some structural rearrangement such as inversion and translocation still happened after the divergence of these strains (Figure 6). In detail, the *Ph. nicotianae* JM01 and *Ph. nicotianae* race 0, *Ph. nicotianae* JM01 and *Ph. nicotianae* race 1, *Ph. nicotianae* JM01 and *Ph. nicotianae* BL shared 12,922, 12,179 and 14,429 blocks, respectively. Between *Ph. nicotianae* JM01 and *Ph. nicotianae* race 0 (Supplementary Table S4), 6269 locations experienced inversion, and 12,663 locations

experienced translocation. Additionally, 3706 inversions and 11,840 translocations were found between *Ph. nicotianae* JM01 and *Ph. nicotianae* race 1 (Supplementary Table S5). In addition, we found that 5182 locations experienced inversion, and 6015 locations experienced translocation between *Ph. nicotianae* JM01 and *Ph. nicotianae* BL (Supplementary Table S6).

Figure 6. Syntenic regions comparison of four *Ph. nicotianae* genomes. The wide horizontal lines in red, blue, and green represent the assembled scaffolds in *Ph. nicotianae* JM01, *Ph. nicotianae* race 0, *Ph. nicotianae* race 1 and *Ph. nicotianae* BL. MUMmer software (Version 3.23) was used to analyze the genome synteny between any two genomes. (**A**) Collinear blocks between *Ph. nicotianae* JM01 and *Ph. nicotianae* race 0. (**B**) Collinear blocks between *Ph. nicotianae* JM01 and *Ph. nicotianae* race 1. (**C**) Collinear blocks between *Ph. nicotianae* JM01 and *Ph. nicotianae* BL.

3. Discussion

3.1. Genome Features

Ph. nicotianae is one of the most significant important plant pathogens [2]. *Phytophthora*-associated diseases are difficult to control through chemical means and other strategies. Moreover, three genomes of *Ph. nicotianae* (*Ph. nicotianae* race 0, race 1 and BL) were available in the NCBI database. Liu et al. (2016). analyzed the RXLR effector and found that the difference in RXLR effector gene numbers between race 0 and race 1 may be related with their pathogenicities in tobacco [26]. However, there were no systematic studies which evaluated the divergence time and virulence mechanisms of *Ph. nicotianae*. Therefore, we assembled the genome of *Ph. nicotianae* JM01 using multiple techniques to understand its pathogenicity mechanisms. The assembled genome of *Ph. nicotianae* JM01 is about 95.32 Mb in size, which is close to the previously published genomes of *Ph. sojae* and *H. arabidopsidis* [16,35], although smaller than that of *Ph. infestans*, with the assembled genome size of 240 Mb. Previous studies showed that the estimated genome size of *Ph. ramorum* is 65 Mb [3], which is smaller than that of *Ph. nicotianae* JM01 [16]. Furthermore, the genome size of *Ph. nicotianae* races 0, *Ph. nicotianae* race 1 and *Ph. nicotianae* BL was about 81 Mb, 71 Mb and 106.75 Mb., respectively [26]. The genome sizes of *Ph. nicotianae* BL and *Ph. nicotianae* JM01 were similar, while they were much larger than those in race 0 and race 1. The assembled genome had a GC content of ~49.02%, which is comparable to those in other species [16]. The genome sizes of species in *Phytophthora* ranged from 52.4 Mb (*Ph. lateralis*) to 236.0 Mb (*Ph. alni var. alni*) [36]. Transposable and repetitive elements can generate mutations and alter the genome size, which can markedly impact the genomic architecture and evolution of species. In this regard, our comparison of repeat elements among the *Phytophthora* species revealed that 21% of the *Ph. nicotianae* JM01 genome comprises repeat elements, which is comparable to that in *Ph. ramorum* (18%) and *Ph. sojae* (28%), while it is lower than that in *Ph. infestans* (74%) [36]. Comparison of the available genome sequences indicated that repeat elements contribute to genome size differences in *Phytophthora* species. We found that prediction methods and databases of repetitive sequences have a great influence on the identification of repetitive sequences. Usually, repetitive sequences are removed before gene prediction. Therefore, the removal of repetitive sequences determines the accuracy of gene prediction. When we predict repeating sequences, we used RepBase in 2005 to perform the repeat element prediction, which may also be a limitation, especially for the gene family comparison. Additionally, we found that the sizes of *Pythium* genomes range from 33.9 Mb to 44.7 Mb [8]. Another reason for the varied genome sizes may be the different sequence approaches. For example, the 64 Mb genome of the *Ph. capsica* was corrected to 100.5 Mb in length by the long-read sequencing method in 2021 [18]. Moreover, the genome completeness evaluated according to core eukaryotic genes showed similarities among those in the published oomycete genomes, representing the accuracy and adequate coverage of the whole genome and gene sets.

3.2. Phylogenetic and Divergence Time Analyses

Phylogenetic analyses among these species suggested that the genera *Phytophthora* and *Pythium* are sister groups, which is agreement with previous studies [6]. In this study, we constructed phylogenic time trees for the brown algae, diatoms, Saprolegniales and Peronosporales based on the currently available genomes. We estimated the divergence time of brown algae, diatoms, and Saprolegniales at 532.18, 220.91 and 102.12 Ma (95% HPD = 466.62–572.21, 197.03–245.15 and 91.93–112.61 Ma), respectively. We also estimated that the ancestor of Peronosporales originated in the Carboniferous, at approximately 348.26 (327.73–385.68) Ma, which is similar to a previously published estimate [37]. In addition, comparative genome analysis showed that these four species shared a total of 664 core genes. According to the annotation results, we found that the core genes are closely related to cellular components, molecular functions, polysaccharide biosynthetic process, etc. In addition, large numbers of long syntenic regions were identified among *Ph. nicotianae* JM01, *Ph. nicotianae* race 0, *Ph. nicotianae* race 1 and *Ph. nicotianae* BL, implied

that they are conserved in genome synteny. The genome sizes varied among these four species. This may be caused by repeat regions or the assembly approaches. The possible reason is that these four strains are physiological races of *Ph. nicotianae*. However, what we should pay attention to is that there are still a lot of translocations, which may be caused by the inconsistency and inversion of the chromosomes. In addition, we noticed that different genome sizes can also lead to the comparison of collinearity modules. For example, there are fewer inversions and translocations between *Ph. nicotianae* BL and *Ph. nicotianae* JM01 than the other two species.

3.3. The Pathogenic Mechanisms of Ph. nicotianae

In plants, the cell wall represents an important barrier to avoid invasion of pathogens and is also an important site of host plant–pathogen interactions. Polysaccharides, such as cellulose, hemicellulose, pectin, along with proteins and lignin are the main content of plant cell walls [38]. Extracellular enzymes produced by the plant pathogenic fungi can degrade the components of the plant cell wall. This not only provides sources of nutrients for their growth, but also promotes their colonization and spreads in tissues. The important role of CWDEs in the pathogenicity of phytopathogenic fungi has attracted the attention of many researchers [8]. In the present study, we identified 44 genes coding carbohydrate enzymes detected in infected hosts unique to the *Ph. nicotianae* JM01 genome, among which some of the AAs, CBMs, CHs and GTs have undergone significant expansion. It has been reported that the *Phytophthora* species harbors the expansion gene datasets encoding secreted effectors and CWDZs, differed from *Pythium* species whose over-represented genes were mainly related to proteolytic degradation and signal transduction [22], which is further supported by the results of our study. In addition, our analysis of the phylogenetic relationships among GH genes indicated that the duplication of these genes may be attributed to their expansion in the *Ph. nicotianae* JM01 genome.

It has been demonstrated that CBEL can reduce or eliminate host defenses in *N. benthamiana* as a cell wall glycoprotein [39]. Previous studies showed that 13, 15, 2 and 2 genes have been identified as CBELs in *Ph. sojae*, *Ph. ramorum*, *H. arabidopsidis*, *A. candida*, *respectively* [35,40]. We found 16 genes encoding CBELs in the *Ph. nicotianae* JM01 genome, which is more than those in other closely related species.

Many effectors have been identified in plant pathogens, and numerous gene-poor regions have harbored rapidly evolving pathogenicity effectors. In the present study, we found a significantly larger number of predicted secreted proteins in the genome of *Ph. nicotianae* JM01 than those in the genomes of *Ph. sojae* and *Ph. ramorum*. The CRN and RXLR are mostly populated genes in *Phytophthora* species [6]. RXLRs have been identified that are abundant in *Phytophthora* species (Jiang et al., 2008). Analysis of the genome sequences of several genomes showed that the *Phytophthora* genome harbors extraordinarily large RXLR families, which were supported by our results. In 2018, Vetukuri et al. (2018) identified pathogenicity-related genes such as RXLR, Crinkle, and Necrosis effectors in *Ph. plurivora* and *Ph. colocasiae* genome [20]. Additionally, 924 candidate RXLR effectors were identified in the *Ph. vignae* genome [21]. In addition, Liu et al. (2016) reported that the difference in RXLR effector gene numbers between *Ph. nicotianae* race 0 and race 1 may be related to their pathogenicities in cultivated tobacco [26]. In our study, we found that some RXLRs had NUDIX, WYL3, MtN3_slv, Acyl_trasnf and MaoC_dehydratas motifs in the *Ph. nicotianae* JM01 genome, while their functions remain to be clarified in pathopoiesis. Brett et al. (2010) proposed that these motifs could aid in the protein entry of these proteins into host cells [41]. Currently, little is known about the function of RXLRs in *Ph. nicotianae* JM01, as few such effectors have been studied in this species. PSE1, a member of the RXLR family of effectors in *Ph. parasitica*, accumulates during appressorium-mediated penetration of host roots and suppresses cell death triggered by both the elicitin cryptogein in *Ph. cryptogea* and AvrPto avirulence proteins in *Ph. syringae* [42]. These observations indicated that *Ph. parasitica* effectors are involved in disease pathogenesis in the pathosystem involving this fungus and *Arabidopsis thaliana* [42]. CRNs, which are

other important and conserved secreted proteins in the oomycete lineage [6,12]. We found a small family of CRN genes in the *Ph. nicotianae* JM01 genome when compared with those in other *Phytophthora species* [8]. However, we found the other elicitors and other effectors showed an expansion in the *Ph. nicotianae* JM01 genome, although these genes are typically similar to those reported in all other species of *Phytophthora*. Combined with the results of phylogenetic analysis, we predicted the expanded pathogenesis-related genes were gained by gene duplication, thereby accelerating the evolutionary adaptation to their hosts. A previous study showed that some pathogenesis-related genes in *S. parasitica* were derived from lateral gene transfer [43]. Different pathogens regulate their methods of infecting host cells through different evolutionary pathways.

Transcription factors existed extensively in different kingdoms. Moreover, it has been found that key elements in transcription factor also tend to be conserved, although there are exceptions, some of which are involved in several spore stages important for pathogenicity and play significant roles in the development of sporangia [4,44]. Pibzp1, which belongs to a bZIP family protein, was previously characterized in *Ph. infestans* and shown a relationship with zoospore motility and infection for their host [45]. In both fungi and plants, bZIPs have been found to regulate responses to the different abiotic stresses like starvation, osmotic and reactive oxygen etc. [4,45]. PsMYB1, an R2R3-type Myb transcription factor from *Ph. sojae*, which are regulated by PsSAK1 and play an important role in lifecycle stages during the infection process [46]. The function of bZIPs performed regulatory effects on *Ph. sojae* via interacting with promoter elements [47]. Meanwhile, some promoter elements have been identified based on the whole genome of *Ph. infestans*, which are closely related to spore formation and infection. The transcription factor Mybal is considered to play an important role in the spore development of oomycetes. This inference is supported by the function of PsMYB1, which is required for sporangium differentiation and the release of zoospores. In *Ph. infestans*, the expression PiMyb2R3, an ortholog of PsMYB1, is reduced during sporulation. We accordingly believe that analyses of the bZIP and Myb transcription factors may improve the understanding of gene regulation in *Ph. nicotianae* JM01. We identified a total of 64 Myb factors in the *Ph. nicotianae* JM01 genome, which is a larger number than those have been found in any other *Phytophthora* genome examined to date. Based on our analyses, we speculated that gene duplication also played important roles in the pathogenesis of *Ph. nicotianae* JM01 and formed a predicted mechanism of *Ph. nicotianae* JM01 (Supplementary Figure S7). The results revealed that gene duplication has contributed to the expansion of CWDEs such as glycoside hydrolases, and effectors such as classical RXLR effectors. Thus, we speculated that the CWDEs in *Ph. nicotiana* JM01 can degrade the content of polysaccharides of tobacco leaves. Subsequently, the effectors, elicitors enter the host and gradually infect the host.

4. Materials and Methods

4.1. Material Culture

Ph. nicotianae (JM01) was isolated from infected tobacco plants in Yunnan Province in 2016 and stored in the microbiology laboratory of the Institute of Tobacco Research Institute of Chinese Academy of Agricultural Sciences. The mycelium was cultivated on V8 medium at 25 °C. Before extracting DNA, the mycelium was collected and then washed with sterilized distilled water for the following use.

4.2. DNA Extraction and Quantification, Genome Sequencing

Genomic DNA was isolated using a standard phenol/chloroform protocol [48]. Breifly, 100 mg samples were quickly ground in liquid nitrogen and then added to a centrifuge tube containing 500 μL lysate. Then, the homogenous material was centrifuged at 12,000× *g* for 5 min. Next, the supernatant was transferred to another tube, and equal volume phenol-chloroformisoamyl alcohol (25:24:1) added. Subsequently, the mixture was slowly homogenized for 2 min and centrifuged at 12,000× *g* for 5 min. Then, the nucleic acids were precipitated by addition of 1000 uL of 75% ethanol twice. Nucleic acids were pelleted

at 12,000× *g* for 10 min. The pellet was dried and was resuspended with 30 μL of ultrapure sterile water and stored at −80 °C. DNA concentration was evaluated using a NanoPhotometer spectrophotometer (IMPLEN, Westlake Village, CA, USA). Then, the high-quality DNA (260/230 ≥ 1.8, 260/280 ≥ 1.8) was used to construct two pair-end DNA libraries with an insert length of 170 bp and 500 bp, respectively, and two mate-pair libraries with an insert length of 2000 bp and 6000 bp, respectively. For pair-end libraries, 1 μg of the DNA was sheared using a Covaris system (Covaris, Inc. Woburn, MA, USA) using the 170 bp or 500 bp program, respectively. Then, the sheared DNA fragments were purified with AMPure XP beads, end-repaired, dA-tailed, and ligated to Illumina universal adapters. After adapter ligation, DNA fragments were further used for size selection. In addition, the sequencing index was added to the former products. For mate-pair libraries, 10 μg (for 2000 bp and 6000 bp size libraries) of genomic DNA was sheared to the desired size by Hydroshear (Digilab, Marlborough, MA, USA), then used for end repair. Fragment sizes were purified from 1% low melting agarose gel and circularized by blunt-end ligation. Similarly, purification, end-repaired, dA-tailed, and ligated to Illumina PE sequencing adapters, along with the sequencing index ligation were performed. Finally, the purified libraries were detected using Agilent Bioanalyzer 2100 (DNA 7500 kit) before sequencing. These libraries were sequenced on the Illumina Hiseq2000 platform (San Diego, CA, USA), generating 100-bp pair-end reads. Trimmomatic [49] was used to remove adapters and sequencing primers from the raw data. Meanwhile, FastQC was used for controlling the quality of the raw data [50].

For SMRT library construction and sequencing, the SMRTbell genomic library was constructed using 5 μg DNA using SMRTbell template kit 1.0 (Menlo Park, CA, USA) according to the manufacturer's instructions and was loaded on the RSII platform (P6C4 chemistry) for sequencing. The raw reads were trimmed for low-quality sequences and the organelle sequences were removed by BLASR [51]. Additionally, chromatin was fixed with formaldehyde, and the fixed DNA was extracted and digested with *Dpn*II for a Hi-C library construction. Subsequently, sticky ends were biotinylated and proximity ligated to form chimeric junctions, which were enriched, and then physically sheared into sequences ranging from 300 to 500 bp. These chimeric fragments representing the original cross-linked long-distance physical interactions were then processed into paired-end sequencing libraries.

4.3. Genome Assembly

For genome assembly, SOAP denovo v2.04 was used to assemble the *Ph. nicotianae* genome with default parameters based on the clean data from two pair-end libraries (170 bp and 500 bp) [52]. Then, the clean reads from 2000 bp and 6000 bp mate-pair libraries, and reads from the Pacbio platform, were used to construct scaffolds using SSPACE [53] and BLASR [51], respectively. Subsequently, the gaps that were produced during the scaffolding construction were closed using PBJelly2 with default parameters [54]. Furthermore, the proximity-guided assembly was also performed based on the Hi-C data used for scaffolding. In detail, the paired-end reads were uniquely mapped to the assembled genome, which then scaffolded using a 3D de novo assembly (3D DNA) pipeline (https://github.com/theaidenlab/3d-dna (accessed on 5 June 2020)) with tuned parameters. The raw data have been deposited in the public database NCBI with project number PRJNA389504. The assembled genome has been uploaded as Dataset 1.

4.4. Repeat Element Prediction and Masking

The assembled genome was first scanned to find repetitive elements to avoid interfering with the results of gene prediction. RepeatMasker v4.0.6 [55] and RepeatModeler [56] were used to search for homologous repeat elements in the RepBase [57] and for de novo repeat element identification in the assembled scaffolds, respectively. In addition, summary data were collected for the repeat types, scores, motifs, and positions of simple sequence repeats (SSRs). Tandem repeat sequences in the genomes of *Ph. nicotianae* were identified

using the Tandem Repeats Finder [58], and repeat elements were masked before conducting gene predictions using RepeatMasker v4.0.6 [55].

4.5. Gene Prediction

Gene prediction for the masked genome of *Ph. nicotianae* JM01 was performed using three approaches: (1) de novo gene prediction was performed using AUGUSTUS 3.0.2 [34]; (2) protein evidence was downloaded from the UniProt/Swiss-Prot protein database [59,60] and the predicted proteins of *Ph. capsici*, *Ph. ramorum*, *Ph. infestans*, *Ph. parasitica*, *Ph. cinnamomic*, *Ph. nicotianae* race 0 and *Ph. sojae* using exonerate [60,61] and (3) EST or cDNA sequences of *Ph. nicotianae* that were downloaded from NCBI were selected to predict gene structures using PASA [6]. When all the prediction results were integrated, we assigned weight scores to different data sets, and then integrated these three prediction results, and finally formed the gene sets of the *Ph. nicotiana* JM01 genome. The transcripts are the direct evidence of coding genes. Therefore, we used the existing EST or cDNA sequence in the database to predict the gene model, and the weight score is 8 points. For Augustus, the weight score is 10 points; the protein-coding region based on the closely related species is given a weighted score of 7 points. Finally, EvidenceModeler (2009) was used to integrate the comprehensive set of genes and gene prediction results [6]. After the prediction of the EVM results and corrections by PASA, the final protein-coding sequence was 23,275. Then, genome assembly and protein completeness were assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) (version 2.0), which contains 4584 conserved core genes in eukaryotes [62].

4.6. Gene Annotation

For gene annotation, the predicted genes were annotated using different databases (Nr, GO, Interpro, Pfam and KEGG). Nr BLAST was performed locally using BlastP with an e-value of e^{-5} [63], and then Gene Ontology (GO) and InterPro were annotated based on the results from Blast2GO [64]. Pfam protein analysis was performed locally using an e-value of e^{-5} [65]. KEGG annotation was analyzed using the KAAS web server (https://www.genome.jp/tools/kaas/ (accessed on 19 September 2020)) [66].

4.7. Gene Expansion and Contraction Analysis

To analyze gene families and gene contents variation in *Ph. nicotianae* JM01 and other closely related species (Supplementary Table S1), we performed gene expansion and contraction analysis using OrthoMCL v4.0 software with an e-value of e^{-5} and proteins in closely related genomes with a similarity greater than 50% were defined as orthologous proteins [67].

4.8. Phylogenetic and Divergence Time Analyses

The phylogenetic tree was constructed using OrthoMCL v4.0 software based on common single-copy genes [67]. For tree construction, *Tetrahymena thermophila* was selected as an outgroup species. MAFFT v7 was used to perform alignments [68]. Finally, protein datasets were used to establish phylogenetic relationships and estimate divergence time using BEAST v1.7.5 [69]. The time constraint was chosen according to the divergence time between *Thraustotheca clavata* and *Ph. infestans* (~363 Ma) [66].

4.9. Identification of the Pathogenic Effectors of Ph. nicotianae JM01

To characterize the pathogenic mechanisms of *Ph. nicotianae* JM01, secreted proteins were identified using SignalP V5.0 [70]. Sequences containing transmembrane domains and organelle-targeting signals were removed from the above predicted secreted proteins. The transmembrane domains of secreted proteins can also be predicted using TMHMM2.0 (TransMembrane prediction using Hidden Markov Models) [71]. Finally, the functions of these proteins were determined using WEGO 2.0 [72]. For subsequent analyses, we analyzed the cell wall-degrading enzymes, elicitors, and effectors of *P. nicotianae* and

available related species. The carbohydrate enzymes (CAZy) in the genomes were predicted using the CAZy Analysis Toolkit with default parameters. For elicitors prediction, we downloaded the known elicitors (INF1s, OPELs, CBELs and PEP-13s), and aggregated them into multiple sequence alignments, and then conducted BLAST analyses. For effectors analysis, hidden Markov models of the SCR, RXLR, ATR13, Elicitin, HLE, PCF, and TOXA effectors were downloaded from Pfam database (http://pfam.xfam.org/ (accessed on 25 January 2021)) and Hmmer software (Hmmsearch, Hinxton, UK) [73] was used to identify candidate genes in the genome that matched this model. We also used the Hmmer package hmmbuild to construct a CRN.hmm model according to known Crinkler genes (CRNs) and then used the hmmsearch package to explore CRN domains in *Ph. nicotianae* JM01.

4.10. Analysis of Basic Region-Leucine Zipper (bZIP) and V-Myb Avian Myeloblastosis Viral Oncogene Homolog (Myb) Sequences

For analysis of bZIP and Myb sequences in *Ph. nicotianae* JM01 genome, tools in Hmmer software [73] were used to format HMM models. Subsequently, the sequence alignment was performed using MUSCLE (www.drive5.com/muscle (accessed on 21 February 2021)) [74]. Then, the sequence logos of bZIP and Myb domains were generated using Hmmer software [73].

4.11. Validation of Ph. nicotianae JM01 Effectors by Constructing Recombinant Agrobacterium Tumefaciens Binary PVX Vectors

Cysteine-rich (SCR) protein is a secreted heparin-binding protein, which is related to cell surface, the extracellular matrix, and biochemical characteristics. To determine the function of effectors, a SCR protein (PnF1) was amplified using the DNA isolated from *Ph. nicotianae* JM01. The PCR product was cloned into the PVX vector pGR107 (Tyler et al., 2006). Then, the accuracy of recombinant binary plasmid was sequenced by Sanger sequencing. After confirmation, the constructs were propagated in *Escherichia coli*, which were cultured in LB medium containing 12.5 mg/mL tetracycline and 50 mg/mL kanamycin. Then, the constructs were selected under the same concentration of tetracycline and kanamycin mentioned above. Subsequently, these constructs were introduced into *A. tumefaciens* strain GV3101 via electroporation [8]. Finally, the recombinant strains of *A. tumefaciens* were infiltrated into *N. benthamiana* leaf using a needle to detect the function of PnF1. Meanwhile, BCL2-Associated X (BAX) (usually used as a positive control for hypersensitive response) and INF1 (which is an extracellular protein secreted by the oomycete *Ph. infestans*), were injected to *N. benthamiana* leaf as a positive control and green fluorescent protein (GFP) and 10 mM $MgCl_2$ were selected as a negative control.

4.12. Genome-Wide Multicollinearity Analysis and Gene Contents Analysis among Four Ph. nicotianae Strains

To infer the evolution of genomic structure in four *Ph. nicotianae* genomes, we used MUMmer software (Version 3.23) to scan the genomes of *Ph. nicotianae* JM 01, race 0, race 1 and BL to identify putative homologous chromosomal regions. Then, we used LASTZ (Version 1.03.54) to compare these regions, and confirmed the arrangement relationships, and found the translocation, inversion, and translocation + inversion regions. In addition, the core genes and specific genes were identified using cd-hit (Version 4.6.1) software. Additionally, the R package was used to draw a Venn diagram [75].

5. Conclusions

Ph. nicotianae, widely distributed in soil and irrigation water, is one of the most significant important pathogens in the Peronosporales. We isolated *Ph. nicotianae* JM01 from tobacco, and then sequenced its whole genome and analyzed its genome features, evolutionary position, and its possible pathogenic mechanisms in this study. Furthermore, we performed a comparative analysis among four *Ph. nicotianae* genomes. The length of the assembled genome of *Ph. nicotianae* JM01 is about 95.32 Mb. Additionally, we identified 19.15% interspersed elements and 0.95% tandem elements across the whole

genome. In addition, a genome-wide phylogenetic tree indicated that *Phytophthora* diverged from *Pythium* by approximately 156.32 Ma. To determine its pathogenic mechanisms, we analyzed a suite of proteins that were involved in plant–pathogen interactions. In addition, transient expression was performed on *N. benthamiana* by infiltrating with *A. tumefaciens* cells containing a cysteine-rich (SCR) protein to detect the function of effector. The genome synteny results showed large amounts of long syntenic regions were shared among *Ph. nicotianae* JM01 and *Ph. nicotianae* race 0, *Ph. nicotianae* JM01 and *Ph. nicotianae* race 1, and *Ph. nicotianae* JM01 and *Ph. nicotianae* BL. Additionally, 680 core common genes shared in these strains. In summary, our analyses of the whole-genome sequence of *Ph. nicotianae* JM01 reinforce previous hypotheses regarding the pathogenic mechanisms of *Phytophthora* and will help us advance the current understanding of infected mechanisms of oomycete plant pathogens.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/plants10081620/s1, Figure S1: BUSCO assessment results of *Ph. nicotianae*, Figure S2: Nr blast results of *Ph. nicotianae* genome, Figure S3: Overview of Go annotation with the *Ph. nicotianae* genome, Figure S4: Barplot of KEGG annotation for the *Ph. nicotianae* genome, Figure S5: Phylogenetic tree for RXLR effectors in *Ph. nicotianae*, Figure S6: Conserved residues in the candidate CRN effector proteins in *Ph. nicotianae*, Figure S7: The predicted mechanism of *Ph. nicotianae*. Table S1: Genome accession numbers and gene names for all proteins included in this study, Table S2: KO annotations results of the *Ph. nicotianae* genome, Table S3: Elicitors in *Ph. nicotianae* and other species, Table S4: Syntenic regions comparison between *Ph. nicotianae* JM01 and *Ph. nicotianae* race 0, Table S5: Syntenic regions comparison between *Ph. nicotianae* JM01 and *Ph. nicotianae* race 1, Table S6: Syntenic regions comparison between *Ph. nicotianae* JM01 and *Ph. nicotianae* BL. Dataset 1: The *Ph. nicotianae* genome.

Author Contributions: F.-L.W.: Conceptualization, writing—review and editing, supervision. C.-S.Z.: conceptualization, writing—review and editing, supervision. X.-L.Y., Z.-F.Z.: writing—original draft. X.-L.Y., Z.-F.Z., F.-Y.K.: formal analysis, Investigation. X.-L.Y., Z.-F.Z., C.-S.Z.: data curation. X.-L.Y., C.-S.Z.: experiment performance. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in this article and the Supplementary Materials.

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

References

1. Jiang, R.H.; Tyler, B.M. Mechanisms and evolution of virulence in oomycetes. *Annu. Rev. Phytopathol.* **2012**, *50*, 295–318. [CrossRef]
2. Leesutthiphonchai, W.; Vu, A.L.; Ah-Fong, A.M.V.; Judelson, H.S. How does phytophthora infestans evade control efforts? Modern insight into the late blight disease. *Phytopathology* **2018**, *108*, 916–924. [CrossRef] [PubMed]
3. Gascuel, Q.; Martinez, Y.; Boniface, M.C.; Vear, F.; Pichon, M.; Godiard, L. The sunflower downy mildew pathogen plasmopara halstedii. *Mol. Plant Pathol.* **2015**, *16*, 109–122. [CrossRef] [PubMed]
4. Judelson, H.S.; Roberts, S. Novel protein kinase induced during sporangial cleavage in the oomycete phytophthora infestans. *Eukaryot. Cell* **2002**, *1*, 687–695. [CrossRef]
5. Judelson, H.S. Metabolic diversity and novelties in the oomycetes. *Annu. Rev. Microbiol.* **2017**, *71*, 21–39. [CrossRef] [PubMed]
6. Haas, B.J.; Kamoun, S.; Zody, M.C.; Jiang, R.H.; Handsaker, R.E.; Cano, L.M.; Grabherr, M.; Kodira, C.D.; Raffaele, S.; Torto-Alalibo, T.; et al. Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. *Nature* **2009**, *461*, 393–398. [CrossRef] [PubMed]
7. Leonelli, L.; Pelton, J.; Schoeffler, A.; Dahlbeck, D.; Berger, J.; Wemmer, D.E.; Staskawicz, B. Structural elucidation and functional characterization of the Hyaloperonospora arabidopsidis effector protein ATR13. *PLoS Pathog.* **2011**, *7*, 15. [CrossRef] [PubMed]

8. Lévesque, C.A.; Brouwer, H.; Cano, L.; Hamilton, J.P.; Holt, C.; Huitema, E.; Raffaele, S.; Robideau, G.P.; Thines, M.; Win, J.; et al. Genome sequence of the necrotrophic plant pathogen pythium ultimum reveals original pathogenicity mechanisms and effector repertoire. *Genome. Biol.* **2010**, *11*, 2010–2011. [CrossRef]

9. Sudhagar, A.; Kumar, G.; El-Matbouli, M. Transcriptome Analysis based on RNA-Seq in understanding pathogenic mechanisms of diseases and the immune system of fish: A comprehensive review. *Int. J. Mol. Sci.* **2018**, *19*, 245. [CrossRef]

10. Chen, X.R.; Xing, Y.P.; Li, Y.P.; Tong, Y.H.; Xu, J.Y. RNA-Seq reveals infection-related gene expression changes in phytophthora capsici. *PLoS ONE* **2013**, *8*, e74588. [CrossRef]

11. Sun, J.; Gao, Z.; Zhang, X.; Zou, X.; Cao, L.; Wang, J. Transcriptome analysis of Phytophthora litchii reveals pathogenicity arsenals and confirms taxonomic status. *PLoS ONE* **2017**, *12*, e0178245. [CrossRef]

12. Gaulin, E.; Pel, M.J.C.; Camborde, L.; San-Clemente, H.; Courbier, S.; Dupouy, M.A.; Lengellé, J.; Veyssiere, M.; Le Ru, A.; Grandjean, F.; et al. Genomics analysis of aphanomyces spp. identifies a new class of oomycete effector associated with host adaptation. *BMC Biol.* **2018**, *16*, 43. [CrossRef] [PubMed]

13. Netea, M.G.; Schlitzer, A.; Placek, K.; Joosten, L.A.B.; Schultze, J.L. Innate and adaptive immune memory: An evolutionary continuum in the host's response to pathogens. *Cell Host Microbe* **2019**, *25*, 13–26. [CrossRef] [PubMed]

14. Rujirawat, T.; Patumcharoenpol, P.; Lohnoo, T.; Yingyong, W.; Lerksuthirat, T.; Tangphatsornruang, S.; Suriyaphol, P.; Grenville-Briggs, L.J.; Garg, G.; Kittichotirat, W.; et al. Draft genome sequence of the pathogenic oomycete pythium insidiosum strain Pi-S, isolated from a patient with pythiosis. *Genome Announc.* **2015**, *3*, e00574-15. [CrossRef] [PubMed]

15. Krajaejun, T.; Kittichotirat, W.; Patumcharoenpol, P.; Rujirawat, T.; Lohnoo, T.; Yingyong, W. Draft genome sequence of the oomycete pythium destruens strain ATCC 64221 from a horse with pythiosis in Australia. *BMC Res. Notes* **2020**, *13*, 329. [CrossRef]

16. Tyler, B.M.; Tripathy, S.; Zhang, X.; Dehal, P.; Jiang, R.H.; Aerts, A.; Arredondo, F.D.; Baxter, L.; Bensasson, D.; Beynon, J.L.; et al. Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* **2006**, *313*, 1261–1266. [CrossRef]

17. Lamour, K.H.; Mudge, J.; Gobena, D.; Hurtado-Gonzales, O.P.; Schmutz, J.; Kuo, A.; Miller, N.A.; Rice, B.J.; Raffaele, S.; Cano, L.M.; et al. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen Phytophthora capsici. *Mol. Plant-Microbe Interact.* **2012**, *25*, 1350–1360. [CrossRef]

18. Shi, J.; Ye, W.; Ma, D.; Yin, J.; Zhang, Z.; Wang, Y.; Qiao, Y. Improved whole genome sequence of phytophthora capsici generated by long-read sequencing. *Mol. Plant-Microbe Interact.* **2021**. [CrossRef] [PubMed]

19. Ali, S.S.; Shao, J.; Lary, D.J.; Kronmiller, B.; Shen, D.; Strem, M.D.; Amoako-Attah, I.; Akrofi, A.Y.; Begoude, B.A.; Ten Hoopen, G.M.; et al. Phytophthora megakarya and P. palmivora, closely related causal agents of cacao black pod rot, underwent increases in genome sizes and gene numbers by different mechanisms. *Genome Biol. Evol.* **2017**, *9*, 536–557. [CrossRef] [PubMed]

20. Vetukuri, R.R.; Tripathy, S.; Malar, C.M.; Panda, A.; Kushwaha, S.K.; Chawade, A.; Andreasson, E.; Grenville-Briggs, L.J.; Whisson, S.C. Draft genome sequence for the tree pathogen phytophthora plurivora. *Genome Biol. Evol.* **2018**, *10*, 2432–2442. [CrossRef]

21. Yang, C.; Lin, L.; Bao, J.; Wang, Z.; Li, Z.; Guo, H.; Lv, L.; Yu, D.; Chen, Q. Genome sequence resource of phytophthora vignae, the causal agent of stem and root rot of cowpea. *Mol. Plant-Microbe Interact.* **2021**. [CrossRef]

22. Adhikari, B.N.; Hamilton, J.P.; Zerillo, M.M.; Tisserat, N.; Lévesque, C.A.; Buell, C.R. Comparative genomics reveals insight into virulence strategies of plant pathogenic oomycetes. *PLoS ONE* **2013**, *8*, e75072. [CrossRef]

23. Wu, J.; Handique, U.; Graham, J.; Johnson, E. Phytophthora nicotianae infection of citrus leaves and host defense activation compared to root infection. *Phytopathology* **2020**, *110*, 1437–1448. [CrossRef]

24. Yang, J.K.; Tong, Z.J.; Fang, D.H.; Chen, X.J.; Zhang, K.Q.; Xiao, B.G. Transcriptomic profile of tobacco in response to phytophthora nicotianae infection. *Sci. Rep.* **2017**, *7*, 401. [CrossRef] [PubMed]

25. Vanegas-Villa, D.M.; Navarro-Alzate, R.A.; Afanador-Kafuri, L.; Gutierrez-Monsalve, J.A.; Morales-Osorio, J.G.; Uribe-Soto, S.I.; Gaviria-Gutierrez, B.M. Effect of interaction between phytophthora nicotianae and meloidogyne spp. on the productivity and quality of tobacco plants (nicotiana tabacum). *Nematology* **2020**, *22*, 1179–1192. [CrossRef]

26. Liu, H.; Ma, X.; Yu, H.; Fang, D.; Li, Y.; Wang, X.; Wang, W.; Dong, Y.; Xiao, B. Genomes and virulence difference between two physiological races of phytophthora nicotianae. *GigaScience* **2016**, *5*, 3. [CrossRef] [PubMed]

27. Drake, K.; Lewis, R.S. An introgressed *Nicotiana rustica* genomic region confers resistance to *Phytophthora nicotianae* in cultivated tobacco. *Crop. Sci.* **2013**, *53*, 1366–1374. [CrossRef]

28. Apple, J. Occurrence of race 1 of phytophthora parasitica var. nicotianae in north carolina and its implications in breeding for disease resistance. *Tob. Sci.* **1967**, *11*, 79–83.

29. Prinsloo, G.; Pauer, G. Die identifikasie van rasse van phytophthora nicotianae B. De Haan var. nicotianae wat in suid-afrika voorkom. *Phytophylactica* **1974**, *6*, 4.

30. McIntyre, J.L.; Taylor, G. Race 3 of phytophthora parasitica var. nicotianae. *Phytopathology* **1978**, *68*, 35–38. [CrossRef]

31. Bowers, J.H.; Locke, J.C. Effect of formulated plant extracts and oils on population density of phytophthora nicotianae in soil and control of phytophthora blight in the greenhouse. *Plant. Dis.* **2004**, *88*, 11–16. [CrossRef] [PubMed]

32. Yuan, X.; Feng, C.; Zhang, Z.; Zhang, C. Complete mitochondrial genome of phytophthora nicotianae and identification of molecular markers for the oomycetes. *Front. Microbiol.* **2017**, *8*, 1484. [CrossRef] [PubMed]

33. Qu, T.; Shao, Y.; Csinos, A.S.; Ji, P. Sensitivity of phytophthora nicotianae from tobacco to fluopicolide, mandipropamid, and oxathiapiprolin. *Plant. Dis.* **2016**, *100*, 2119–2125. [CrossRef] [PubMed]

34. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **2006**, *34*, W435–W439. [CrossRef]

35. Baxter, L.; Tripathy, S.; Ishaque, N.; Boot, N.; Cabral, A.; Kemen, E.; Thines, M.; Ah-Fong, A.; Anderson, R.; Badejoko, W.; et al. Signatures of adaptation to obligate biotrophy in the hyaloperonospora arabidopsidis genome. *Science* **2010**, *330*, 1549–1551. [CrossRef]

36. McGowan, J.; Fitzpatrick, D.A. Recent advances in oomycete genomics. *Adv. Genet.* **2020**, *105*, 175–228. [CrossRef]

37. Matari, N.H.; Blair, J.E. A multilocus timescale for oomycete evolution estimated under three distinct molecular clock models. *BMC Evol. Biol.* **2014**, *14*, 1471–2148. [CrossRef]

38. Höfte, H.; Voxeur, A. Plant cell walls. *Curr. Biol.* **2017**, *27*, R865–R870. [CrossRef]

39. Mateos, F.V.; Rickauer, M.; Esquerré-Tugayé, M.T. Cloning and characterization of a cDNA encoding an elicitor of phytophthora parasitica var. nicotianae that shows cellulose-binding and lectin-like activities. *Mol. Plant-Microbe Interact.* **1997**, *10*, 1045–1053. [CrossRef]

40. Links, M.G.; Holub, E.; Jiang, R.H.; Sharpe, A.G.; Hegedus, D.; Beynon, E.; Sillito, D.; Clarke, W.E.; Uzuhashi, S.; Borhan, M.H. De novo sequence assembly of albugo candida reveals a small genome relative to other biotrophic oomycetes. *BMC Genom.* **2011**, *12*, 1471–2164. [CrossRef]

41. Kale, S.D.; Gu, B.; Capelluto, D.G.; Dou, D.; Feldman, E.; Rumore, A.; Arredondo, F.D.; Hanlon, R.; Fudal, I.; Rouxel, T.; et al. External lipid PI3P mediates entry of eukaryotic pathogen effectors into plant and animal host cells. *Cell* **2010**, *142*, 284–295. [CrossRef]

42. Evangelisti, E.; Govetto, B.; Minet-Kebdani, N.; Kuhn, M.L.; Attard, A.; Ponchet, M.; Panabières, F.; Gourgues, M. The phytophthora parasitica RXLR effector penetration-specific effector 1 favours arabidopsis thaliana infection by interfering with auxin physiology. *New Phytol.* **2013**, *199*, 476–489. [CrossRef]

43. Jiang, R.H.; de Bruijn, I.; Haas, B.J.; Belmonte, R.; Löbach, L.; Christie, J.; van den Ackerveken, G.; Bottin, A.; Bulone, V.; Díaz-Moreno, S.M.; et al. Distinctive expansion of potential virulence genes in the genome of the oomycete fish pathogen saprolegnia parasitica. *PLoS Genet.* **2013**, *9*, e1003272. [CrossRef] [PubMed]

44. Tani, S.; Yatzkan, E.; Judelson, H.S. Multiple pathways regulate the induction of genes during zoosporogenesis in phytophthora infestans. *Mol. Plant-Microbe Interact.* **2004**, *17*, 330–337. [CrossRef] [PubMed]

45. Blanco, F.A.; Judelson, H.S. A bZIP transcription factor from phytophthora interacts with a protein kinase and is required for zoospore motility and plant infection. *Mol. Microbiol.* **2005**, *56*, 638–648. [CrossRef] [PubMed]

46. Zhang, M.; Lu, J.; Tao, K.; Ye, W.; Li, A.; Liu, X.; Kong, L.; Dong, S.; Zheng, X.; Wang, Y. A Myb transcription factor of Phytophthora sojae, regulated by MAP kinase PsSAK1, is required for zoospore development. *PLoS ONE* **2012**, *7*, e40246. [CrossRef] [PubMed]

47. Ye, W.; Wang, Y.; Dong, S.; Tyler, B.M. Phylogenetic and transcriptional analysis of an expanded bZIP transcription factor family in phytophthora sojae. *BMC Genom.* **2013**, *14*, 1471–2164. [CrossRef] [PubMed]

48. Kerkhof, L.; Ward, B.B. Comparison of nucleic acid hybridization and fluorometry for measurement of the relationship between RNA/DNA ratio and growth rate in a marine bacterium. *Appl. Environ. Microbiol.* **1993**, *59*, 1303–1309. [CrossRef]

49. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

50. de Sena Brandine, G.; Smith, A.D. Falco: High-speed FastQC emulation for quality control of sequencing data. *F1000Research* **2019**, *8*, 1874. [CrossRef]

51. Chaisson, M.J.; Tesler, G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinform.* **2012**, *13*, 238. [CrossRef]

52. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 1–18. [CrossRef]

53. Boetzer, M.; Henkel, C.V.; Jansen, H.J.; Butler, D.; Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **2011**, *27*, 578–579. [CrossRef]

54. English, A.C.; Richards, S.; Han, Y.; Wang, M.; Vee, V.; Qu, J.; Qin, X.; Muzny, D.M.; Reid, J.G.; Worley, K.C.; et al. Mind the gap: Upgrading genomes with pacific biosciences RS long-read sequencing technology. *PLoS ONE* **2012**, *7*, e47768. [CrossRef]

55. Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* **2004**, *5*, 4.10. 1–4.10. 14. [CrossRef] [PubMed]

56. 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]

57. Jurka, J.; Kapitonov, V.V.; Pavlicek, A.; Klonowski, P.; Kohany, O.; Walichiewicz, J. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **2005**, *110*, 462–467. [CrossRef]

58. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef] [PubMed]

59. Apweiler, R.; Bairoch, A.; Wu, C. The universal protein resource (UniProt). *Nucleic Acids Res.* **2008**, *36*, D190–D195. [CrossRef]

60. Bairoch, A.; Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL. *Nucleic Acids Res.* **2000**, *28*, 45–48. [CrossRef] [PubMed]

61. Slater, G.S.; Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **2005**, *6*, 1471–2105. [CrossRef]

62. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [CrossRef]
63. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
64. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [CrossRef]
65. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]
66. Moriya, Y.; Itoh, M.; Okuda, S.; Yoshizawa, A.C.; Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **2007**, *35*, W182–W185. [CrossRef] [PubMed]
67. Li, L.; Stoeckert, C.J., Jr.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **2003**, *13*, 2178–2189. [CrossRef] [PubMed]
68. 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]
69. Drummond, A.J.; Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **2007**, *7*, 1471–2148. [CrossRef] [PubMed]
70. Torto, T.A.; Li, S.; Styer, A.; Huitema, E.; Testa, A.; Gow, N.A.; van West, P.; Kamoun, S. EST mining and functional expression assays identify extracellular effector proteins from the plant pathogen phytophthora. *Genome Res.* **2003**, *13*, 1675–1685. [CrossRef]
71. Sonnhammer, E.L.; von Heijne, G.; Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1998**, *6*, 175–182.
72. Ye, J.; Zhang, Y.; Cui, H.; Liu, J.; Wu, Y.; Cheng, Y.; Xu, H.; Huang, X.; Li, S.; Zhou, A.; et al. WEGO 2.0: A web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* **2018**, *46*, W71–W75. [CrossRef] [PubMed]
73. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, 18. [CrossRef] [PubMed]
74. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
75. Chen, H.; Boutros, P.C. VennDiagram: A package for the generation of highly-customizable venn and euler diagrams in R. *BMC Bioinform.* **2011**, *12*, 35. [CrossRef] [PubMed]