scientific data

OPEN

DATA DESCRIPTOR

Chromosome-scale assembly and annotation of *Phytophthora capsici* isolate BYA5

Yiman Wan[1,2,7], Fuqiang Zhu[2,3,7], Meng Zhang[2], Xiaohui Li[2,4], Chunxue Xie[2,5], Shenglan Gao[2,6], Qunqing Wang[1]✉ & Yuan Chen[2,6]✉

Phytophthora capsici **is a globally widespread oomycete pathogen that causes severe damage to diverse host plants, posing major agricultural challenges. Here, we present a high-quality, chromosome-level genome assembly of *P. capsici* isolate BYA5 using PacBio HiFi, ultralong Oxford Nanopore, and Hi-C sequencing data. The assembled genome spans 83.96 Mb and comprises 18 chromosomes, capturing all centromeres and most telomeres (32/36). The assembled genome exhibits a contiguous N50 of 3.84 Mb, representing the first chromosome-level genome sequence of a pathogenic *P. capsici*. A total of 15,688 protein-coding genes were annotated, including 337 RxLRs and 115 Crinkler (CRN) effectors. This high-quality reference genome provides a valuable resource for advancing our understanding of oomycete pathobiology and evolution, facilitating further research on virulence mechanisms and host-pathogen interactions.**

Background & Summary

Pepper (*Capsicum annuum* L.) is an economically important crop worldwide, but its production is seriously threatened by *Phytophthora capsici*, a highly destructive oomycete pathogen[1]. This disease affects all parts of the plant, including roots, leaves, and fruits, and often causes widespread outbreaks and significant yield losses in key pepper-producing area[2]. Pathogen-derived effectors are key virulence factors that manipulate host physiology to facilitate infection[3]. Effectors that are specifically detected by host resistance (R) proteins are classified as avirulence (Avr) effectors[4]. Among oomycetes, the most extensively studied Avr effectors are members of the RxLR (arginine–any amino acid–leucine–arginine) family. These effectors are broadly distributed across oomycete species and are particularly prevalent in the *Phytophthora* genus[5,6]. During the protracted evolutionary arms race between hosts and pathogens, these virulence proteins undergo rapid diversification to evade detection by plant immune surveillance systems.

For many oomycetes, high-quality assemblies based on long-read sequencing have now been achieved, with some reaching telomere-to-telomere (T2T) levels, such as *Phytophthora sojae*[7], or near-chromosome-level assemblies, as seen in *Phytophthora plurivora*[8], *Phytophthora infestans*[9] and *Phytophthora agathidicida*[10]. These high-quality assemblies have greatly advanced research on genome evolution, structural variation, and pathogenicity in *Phytophthora*.

To date, 18 genome assemblies of *P. capsici* isolates have been made publicly available in genome databases, generated using various next-generation sequencing platforms. Although a chromosome-level assembly (accession number GCA_030324255.1) has recently become available, most existing assemblies remain at the scaffold or contigs level, and the overall genome quality is still limited due to substantial genetic divergence among physiological races and high chromosomal variability across oomycetes. Several isolates from the

[1]College of Plant Protection, Shandong Agricultural University, Tai'an, 271018, Shandong, China. [2]State Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, Shandong, 261325, China. [3]State Key Laboratory of Wheat Improvement, College of Life Sciences, Shandong Agricultural University, Tai'an, Shandong, 271018, China. [4]College of Life Sciences, National Key Laboratory of Crop Improvement for Stress Tolerance and Production, Northwest Agriculture and Forestry University, Yangling, Shanxi, 712100, China. [5]State Key Laboratory of Wheat Improvement, College of Agronomy, Shandong Agricultural University, Tai'an, Shandong, 271018, China. [6]Southwest United Graduate School, Kunming, 650092, China. [7]These authors contributed equally: Yiman Wan, Fuqiang Zhu. ✉e-mail: wangqunqing@163.com; yuan.chen@pku-iaas.edu.cn

Statistics	illumina	PacBio HiFi	Hi-C	Ultra-long
Library size (bp)	350	15,000	350	50,000
Raw data (Gb)	5.70	6.60	10.16	7.70
N50 (bp)	150	11,210	150	50,000
Longest reads (bp)	150	38,755	150	824,583
Mean read length (bp)	150	10,934	150	49,024
Coverage (X)	68.01	78.75	121	91.87

Table 1. The statistics for sequencing data output of *P. capsici* isolate BYA5.

Genome assembly Value	
Number of contigs	18
Assembly size (Mb)	83.96
Conting N50 (bp)	3,841,935
Scaffold N50 (bp)	3,879,088
Longest Scaffold (bp)	8,587,003
Number of genes	15,688
BUSCO completeness (%)	94.10%
Repetitive sequences (% of genome)	
SINEs (bp)	354,581 (0.43%)
LTR (bp)	20,654,979 (25.07%)
Other (bp)	728,720 (0.88%)
DNA transposons (bp)	8,672,597 (10.53%)
non-coding RNAs sequences (% of genome)	
miRNA (bp)	4,973 (0.006%)
tRNA (bp)	552,322 (0.670%)
rRNA (bp)	433,753 (0.526%)
snRNA (bp)	59,706 (0.072%)
Gene annotation	
Number of genes	15,688
Average transcript length (bp)	1,623.09
Average CDS length (bp)	1,434.25
BUSCO completeness (%)	94.10%

Table 2. Statistics for genome assembly and annotation of *P. capsici* isolate BYA5.

United States[11] and Mexico[12] were sequenced using second-generation technologies, expanding genomic resources for this important pathogen. In addition, nine isolates have been sequenced with long-read platforms such as Pacific Biosciences (PacBio)[13,14] and Oxford Nanopore Technologies (ONT). Representative isolates include LT1534 (Pc331)[15], its recurrent backcross parent Pc285[15], LT263[16], SD33[17], and strain 05-06[18]. While 05-06 was sequenced using both PacBio and ONT, the others were sequenced with ONT alone.

Nevertheless, T2T or comprehensive chromosome-level assemblies remain rare, and large-scale comparative genomic analyses across oomycetes are still lacking.

A chromosome-level genome sequence of *P. capsici* is essential for elucidating its pathogenicity, host adaptation, and virulence strategies. Such a resource also deepens our understanding of oomycete evolution and genome architecture, providing a valuable foundation for studies on plant–pathogen interactions.

In this study, we used the *P. capsici* isolate BYA5[19], a highly virulent strain originally obtained from a diseased pepper leaf lesion in Gansu Province, China, in 2011, to generate a chromosome-level reference genome and update its genome annotation based on the improved assembly. We sequenced the genome of *P. capsici* BYA5 using PacBio single-molecule real-time (SMRT) sequencing to obtain high-fidelity (HiFi) long reads, complemented with Illumina short-read sequencing data.In total, we obtained 6.60 Gb of PacBio long-read sequencing data (~78.75 × coverage), 7.70 Gb of ultra-long sequencing data through Nanopore sequencing technology (~91.87 × coverage), and 5.97 Gb of Illumina short-read sequencing data (~68.01 × coverage) (Table 1). The initial genome assembly was generated using the Hifiasm[20] assembler with PacBio HiFi reads. Following the removal of mitochondrial[21], heterozygosity and repetitive sequences, the primary contigs were anchored into pseudochromosomes utilizing Hi-C data with Allhic[22] and juicebox[23]. Identified gaps were filled using ONT-ultra-long reads data with Tgs-gapcloser[24]. The final genome assembly comprised 83.96 Mb, with a contig N50 of 3.84 Mb and a scafold N50 of 3.88 Mb (Table 2).

The majority of the genome assembly (97.48%) was successfully anchored to 18 pseudochromosomes (Fig. 1a). Notably, the assembly contains only a single gap and 32 telomeres were identified across these 18 chromosomes based on the presence of the telomeric repeat sequence (TTTAGGG) (Table 3). The quality of

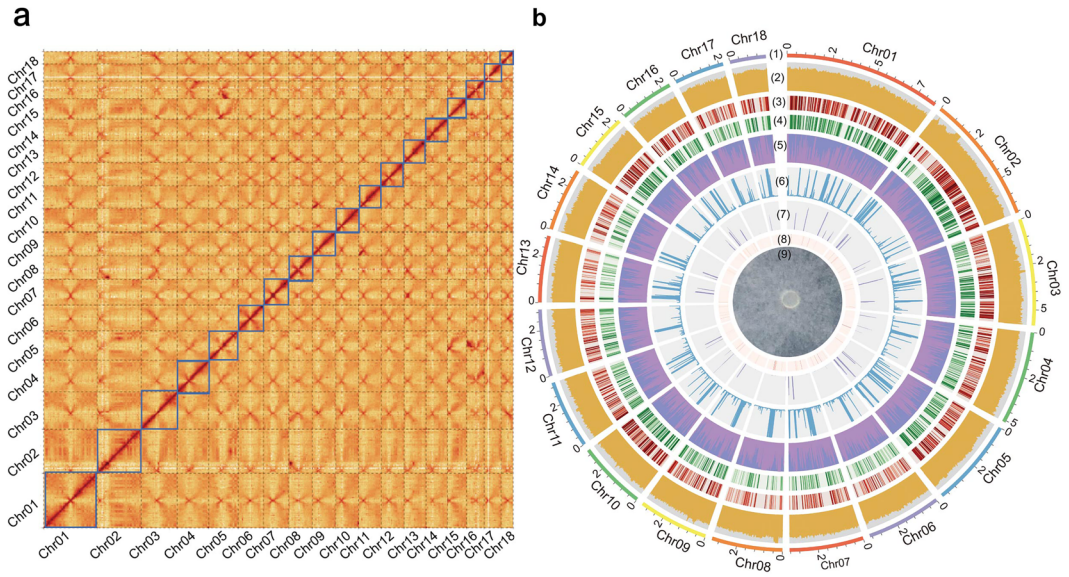

Fig. 1 Overview of a chromosome-level genome assembly and annotation of the *P. capsici* isolate BYA5. (**a**) Chromosome heatmaps of Hi-C data of *P. capsici* isolate BYA5 genome. (**b**) Circos plot of BYA5 genome. Track (1) to (9): (1) chromosome sizes, (2) GC content, (3) gene density grey: low, red: high, (4) exon density green: high, (5) histogram of repetitive DNA density, (6) Simple repeat density, (7) tRNA density, (8) RxLR effector density and (9) Colony morphology of BYA5 strain is photographed after 5-day incubation at 25 °C. The densities are calculated in 20 kb windows (2) to (8). Chr: chromosome. The unit of chromosome length is Mb (1).

Genome	Sequences Length	Gap number	Prediction of telomeres number	Gene number
Chr01	8,587,003	—	2	1,096
Chr02	7,075,242	—	1	1,613
Chr03	5,846,668	1	2	1,231
Chr04	5,104,336	—	2	1,051
Chr05	4,866,279	—	2	1,026
Chr06	4,112,719	—	2	656
Chr07	4,058,032	—	2	860
Chr08	3,879,088	—	2	496
Chr09	3,841,935	—	2	736
Chr10	3,761,117	—	2	904
Chr11	3,748,853	—	2	781
Chr12	3,661,347	—	2	771
Chr13	3,647,817	—	2	441
Chr14	3,471,967	—	2	582
Chr15	3,149,453	—	1	738
Chr16	2,979,603	—	1	580
Chr17	2,712,905	—	1	827
Chr18	1,967,921	—	2	412
total	82,358,918	1	32	15,688
Unplaced	5,886,633	—	—	273

Table 3. Detailed summary of assembled chromosomes.

the assembly was further validated by a Benchmarking Universal Single-Copy Orthologs (BUSCO)[25] score of 94.10% using the eukaryota_odb10 database, and a consensus quality value (QV) score of 42.68, all refecting the high accuracy and completeness of the genome assembly.

This assembly genome captured all centromeres on 18 chromosomes. A total of 32 telomeres were identified, missing 4 telomeres including the 5′ end of chromosomes 17, and the 3′ end of chromosomes 2, 15, and 16 (Fig. 2). Compared to LT263 strain genome assembly (78.80 Mb, N50 of 1.60 Mb; GCA_030324255.1), our assembly is larger (83.96 Mb) and exhibits greater continuity, with a contig N50 of 3.84 Mb (Table 4). Through the analysis of collinearity, we found that BYA5 and the genome of LT263 have a relatively good collinearity

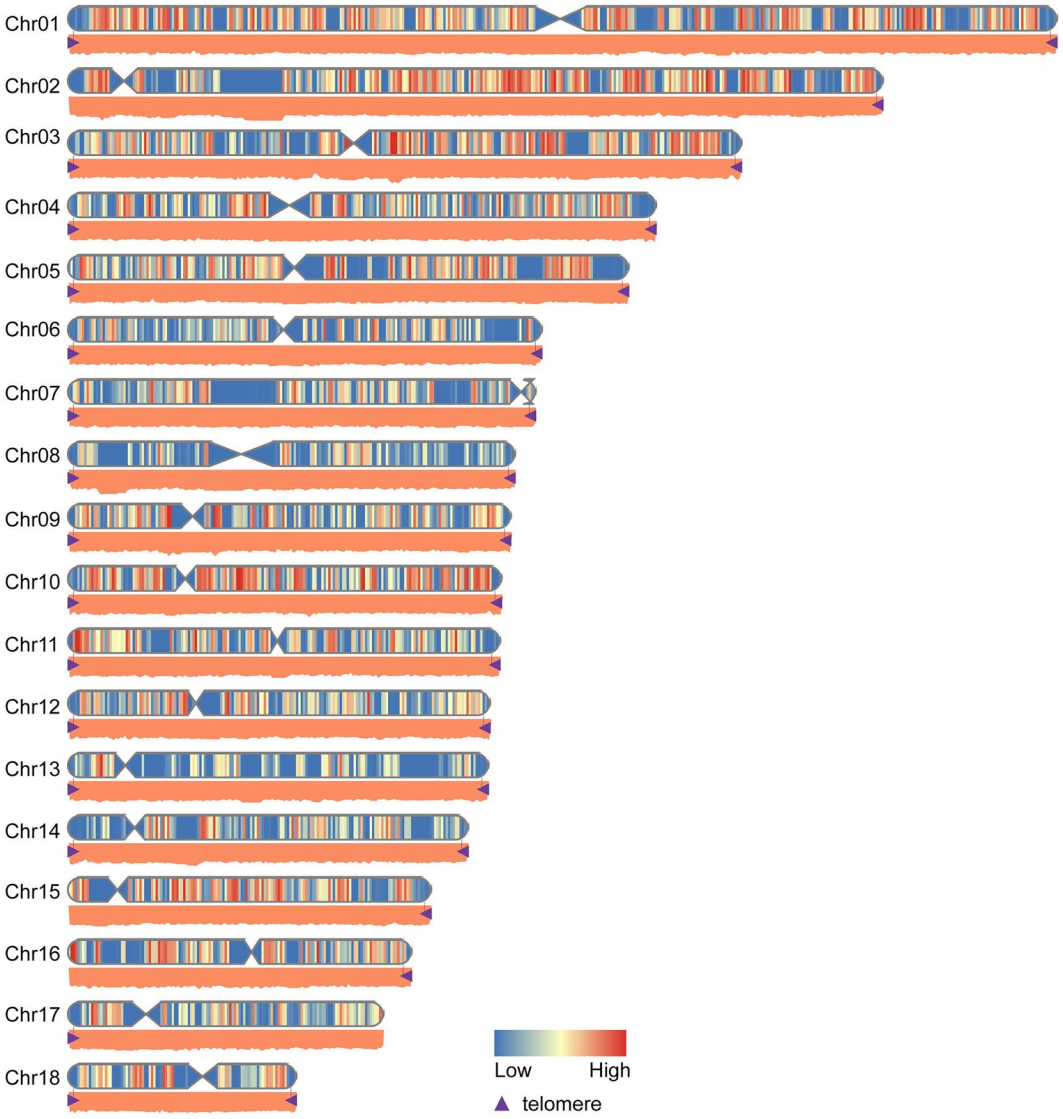

Fig. 2 Localization of telomeres and centromeres and the GC content pattern in the assembled genome of *P. capsici* isolate BYA5. The density of the protein-coding genes was calculated in 5 Mb windows. GC content was shown as graph below each chromosome, purple arrowheads indicate the telomeres position. One centromer is localized within each chromosome.

overall. Chromosomes 1 and 2 of BYA5 respectively fused the information of chromosomes 14 and 13, and chromosomes 17 and 18 of LT263 (Fig. 3a). BYA5 and the genome of other *Phytophthora* also have a relatively good collinearity overall (Fig. 3b).

For annotation of the protein-coding genes, three different sources of evidence were integrated, including *de novo* prediction[26], *Phytophthora* homologous proteins and RNA-seq data. A total of 15,688 protein-coding genes were identified in our assembly, and 91.05% of the genes could be predicted for their functions (Fig. 1b; Table 2). In our assembly, Repetitive elements (REs) accounted for 39.71%. The most abundant REs were Long terminal repeat (LTR) retrotransposons (~25.07% of REs).

Effectors play a critical role in the early stages of *Phytophthora* infection by suppressing host immunity, inhibiting enzymes, and altering cell wall integrity. They manipulate plant structure and metabolism to compromise host defenses, there by facilitating pathogen invasion and colonization[27–30]. Notably, a repertoire of highly diversified RxLR effectors is translocated into host cells, where they orchestrate multilayered immune suppression to promote infection. To identify RxLR effectors in *P. capsici*, we applied a previously established RxLR prediction pipeline to screen the entire sequenced genome of *P. capsici*[31]. This analysis revealed 337 RxLRs (Fig. 4a and Table S1) and 115 Crinkler (CRN) effectors (Fig. 4b and Table S1) in *P. capsici* BYA5. In summary, this chromosome-level genome assembly of *P. capsici* provides a valuable resource for future studies, offering improved continuity and accuracy. It enhances understanding of *Phytophthora* biology and the evolutionary dynamics of oomycetes, particularly in the pepper-*P. capsici* pathosystem.

Statistics	LT263	This study	Difference (%)
Number of contigs	117	18	−0.846
Assembly size (Mb)	78.80	83.96	+0.065
Conting N50 (bp)	1,600,000	3,841,935	+1.424
Scaffold N50 (bp)	4,200,000	3,879,088	−0.0764
Longest Scaffold (bp)	6,028,800	8,587,003	+0.4243
Number of genes	21,752	15,688	−0.3865

Table 4. Comparison of the genome assembly of *P. capsici* isolate BYA5 with that of isolate LT263.

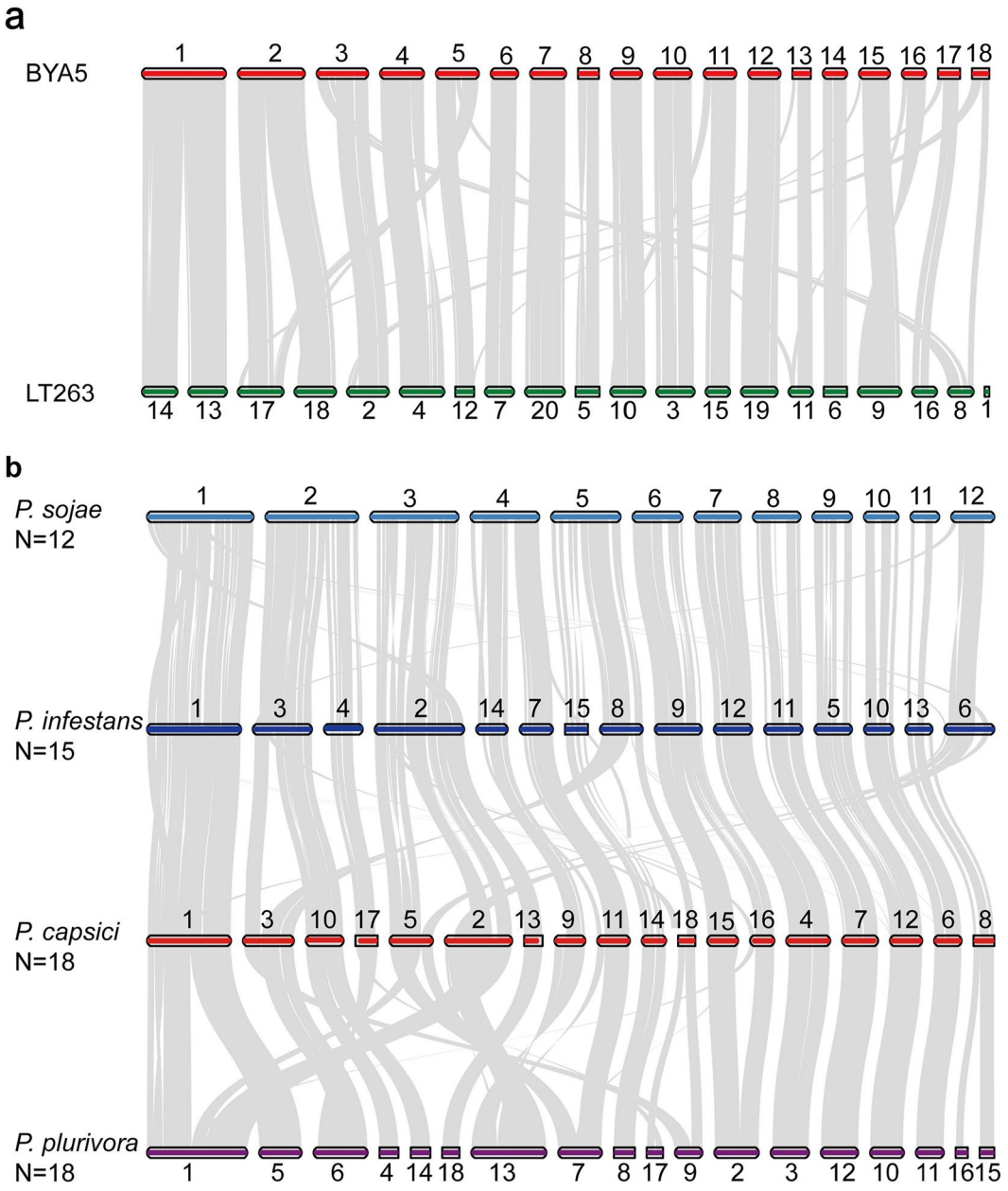

Fig. 3 Chromosome-level synteny analysis between *P. capsici* and three representative *Phytophthora* species. (**a**) Synteny analysis was performed between *P. capsici* isolate BYA5 and LT263. (**b**) Synteny analysis between *P. capsici* isolate BYA5 and other *Phytophthora* species.

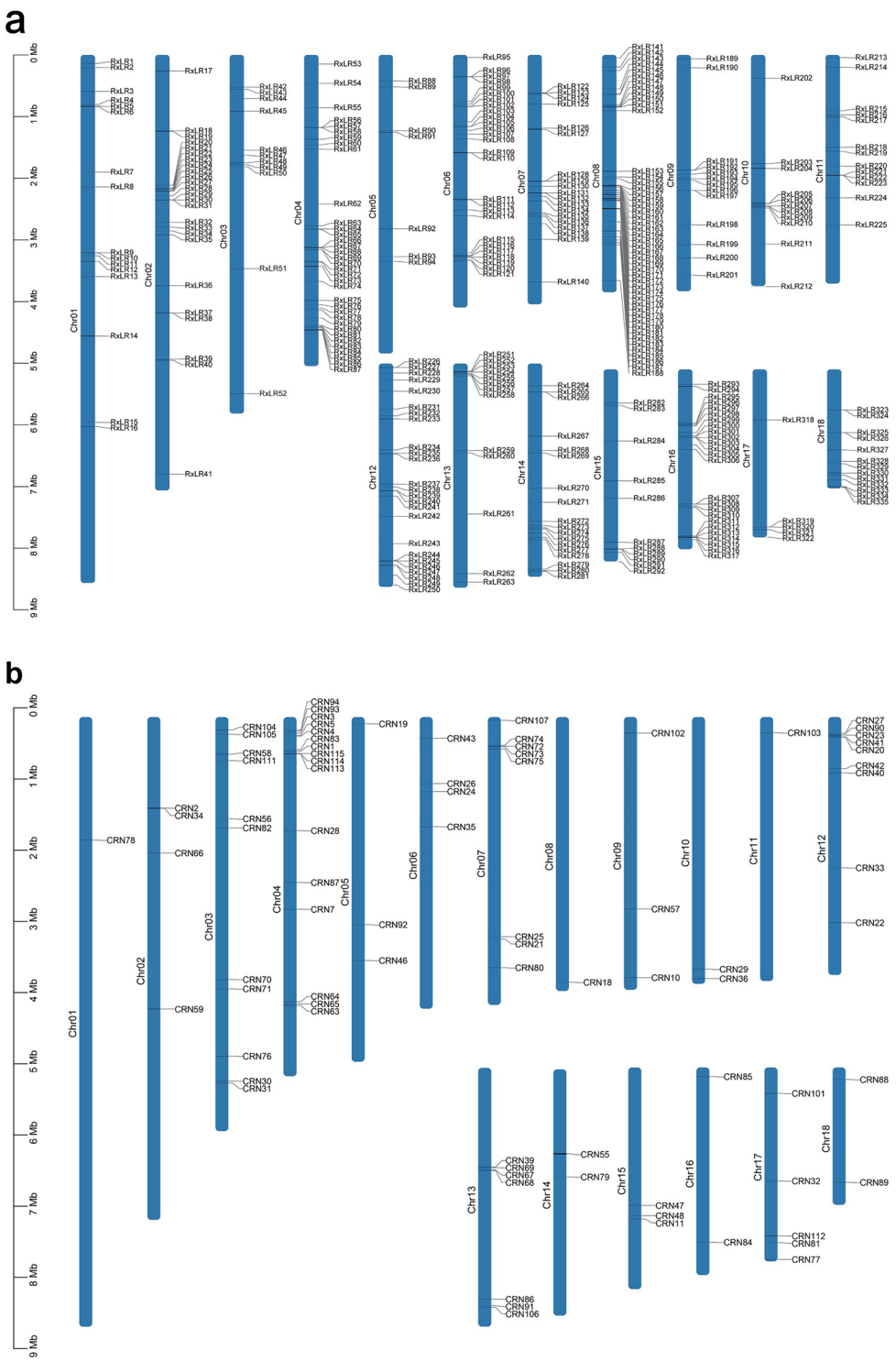

Fig. 4 Distribution of RxLR and CRN effectors in the *P. capsici* isolate BYA5. (**a**) The count of RxLR effectors on each chromosome. (**b**) The count of CRN effectors on each chromosome.

Methods

Sampling and DNA extraction. *P. capsici* isolate BYA5[19] was inoculated on 10% V8 plates and cultured in the dark at 25 °C. After 5 days of culture, 10 mycelial blocks of 4 × 4 mm were cultured in V8 liquid medium at 25 °C in the dark for 5 days. Then, the mycelium was filtered through sterile gauze, the excess moisture of the mycelium was dried with absorbent paper, and the sample was stored in liquid nitrogen (~3 g per sample). Genomic DNA extracted from vegetative mycelia using cetyltrimethylammonium bromide (CTAB) protocol was used for genome sequencing[32]. The quality of the DNA was checked using an Advanced Analytical Technologies, Inc. (AATI) Fragment Analyzer™ and a pulse field gel electrophoresis apparatus (Bio-Rad) at Novogene Bioinformatics Technology Co., Ltd. (Tianjin, China).

Illumina, PacBio and Hi-C sequencing. High molecular weight (HMW) DNA was used for single-molecule sequencing, whole-genome sequencing, and Hi-C sequencing at Novogene Bioinformatics Technology Co., Ltd. (Tianjin, China). For genome surveys, approximately 5.70 Gb (68.01 × coverage) of 150 bp paired-end reads were sequenced on an Illumina Hiseq NovaSeq™ X Plus. For PacBio HiFi sequencing, 15–20 kb libraries were constructed using the SMRTbell Preparation kit 3.0 according to the manufacturer's instructions. The genomic DNA was fragmented using a Covaris G-Tube device, and the size and concentration were measured with an Advanced Analytical Technologies, Inc. (AATI) Fragment Analyzer™. DNA damage repair, end repair, A-tailing and hairpin adapter ligation were performed using the Template Prep Kit. The library was then treated with nuclease using Revio Palymerase kit. Library quality was detected by BluePippin System and set size cut-off threshold depending on the goals of the project. Then, AMPure PB Beads were used to concentrate and purify SMRTbell templates after size selection. For ONT sequencing, Ultra-high molecular weight (uHMW) genomic DNA (gDNA) was extracted to generate ultra-long read libraries (ultra-long read length N50s > 50 kb) library preparation was performed using the Ultra-Long DNA Sequencing Kit (SQK-ULK114; Oxford Nanopore Technologies) following the manufacturer's protocol. uHMW gDNA was tagmented using a diluted fragmentation mix. Rapid Adapters were then attached to the DNA ends. The sample was subsequently purified by DNA precipitation, and the final library was eluted overnight. Library quality was detected by BluePippin System and set size cut-off threshold depending on the goals of the project. Then, start a sequencing run using the MinKNOW software on the PromethION sequencing platform, which will collect raw data from the device and convert it into basecalled reads. For Hi-C sequencing, libraries were prepared from mycelium of BYA5 by cross-linked using a 4% formaldehyde solution to capture interactions, following a published protocol[33]. After enzymatic digestion with 400 units of DpnII, the cross-linked DNA fragments were blunt-end ligated after biotin-14-dCTP was added to the DNA ends. Biotin was removed from non-ligated fragment ends using T4 DNA polymerase. After re-ligation, reverse cross-linking, and purifcation, the chromatin DNA was sheared to a size of 200–600 bp by sonication. The biotin-labeled fragments were then enriched using streptavidin C1 magnetic beads. After adding A-tails to the fragment ends and following ligation by the illumina paired-end (PE) sequencing adapters, Hi-C sequencing libraries were amplified by PCR (12–14 cycles) and sequenced on Illumina PE150. 10.16 Gb paired-end sequencing data (150-bp length) were generated.

Genome survey. Raw next-generation sequencing data were pre-processed using redup v2 software to filter out adaptor sequences, low-quality reads, and short reads with default parameters. Next, the frequency distribution of the depth of clean data with 17-mers (kmer size = 17) was computed using Jellyfish v2.2.7[34]. Genome size was estimated using the following formula: G = K-num / K-depth, where G represents the genome size, K-num is the total number of 3,955,906,298-mers, and K-depth denotes the 46-mer depth. The heterozygosity and repeat content of the genome were estimated by combining simulation data results with varying heterozygosity levels and the frequency peak distribution of 17-mers. The genome size was determined to be approximately 83.81 Mb, with a repeat content ratio of 44.41%.

Genome assembly. For genome assembly, HiFi circular consensus sequencing (CCS) reads were generated from raw subread data using the CCS software (https://github.com/PacificBiosciences/ccs) with stringent quality control parameters (min-rq = 0.99). The filtered HiFi long reads were used for de novo assembly into contigs using hifiasm v0.19.9-r616[20]. To anchor the contigs to chromosomes, Hi-C sequencing data were mapped onto the assembled contigs using the ALLHIC v0.9.8[22] and JuiceBox[23] pipeline, where chromatin interaction signals were processed with enzymatic digestion site parameter "enz = DpnII" and clustering parameter "CLUSTER = N" to orient and order contigs into chromosomal scaffolds. The Hi-C sequencing data were used to anchor all 99 contigs using ALLHIC v0.9.8 pipeline (allhic extract group.clean.bam group.fasta–RE GATC allhic partition–pairsfile group.clean.pairs.txt–contigfile group.clean.counts_GATC.txt-K 19–minREs 50–maxlinkdensity 3–NonInformativeRabio 0), resulting in 18 scafolds, which were further refned using JuiceBox v1.11.08[23]. Finally, the ONT ultra-long reads data were filled in the gaps existing in the scaffolds using Novogene's proprietary GAP-repair technology (Patent No. CN202310944042.X) and TGS-GapCloser v1.2.0[24] software, and then corrected by Racon (v1.4.20). The final genome assembly (82.36 Mb) only 1 gap (Fig. 1a, b) with a contig N50 of 3.84 Mb.

RNA sequencing and analysis. Total RNA was extracted from the zoospore and 12 h, 24 h, 48 h of infected pepper tissue of *P. capsici* using Trizol (Termal Fisher) agents following manufacturer recommendation protocol. The Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) was used to evaluate the total RNA integrity. The total RNA used for library preparation first enriched the mRNA with polyA tail through Oligo (dT) magnetic beads. The mRNA was used to construct sequencing libraries with the Illumina TruSeq Transcriptome Kit (Illumina, CA), with an insert size of 370–420 bp. Sequencing was performed on the Illumina NovaSeq. 6000 platform at Gene Denovo Biotechnology (Guangzhou, China), generating 150 bp paired-end reads. RNA-seq data was checked for quality using fastp v0.23.2[35], mapped to *P. capsici* genome assembly using Hisat2 v.2.1[36].

Gene model and function annotations. To annotate the assembly, three approaches were combined: ab initio prediction, homology-based protein predictions, and RNA-seq evidence. During the ab initio prediction process, the repetitive sequences dispersed in the genome were de novo predicted and compiled by RepeatModeler v2.0.3[37] (http://www.repeatmasker.org/) (parameters: -engine ncbi -pa 30 -LTRStruct). For further analyses, RepeatMasker v4.1.2.p1 (http://www.repeatmasker.org/) (parameters: -nolow -no_is -norna -pa 30) was used to extract and sofmask the repetitive elements. For gene predication based on ab initio, Augustus v3.5 (parameters:–species = pasa1–uniqueGeneId = TRUE–noInFrameStop = TRUE–GFF3 = on–genemodel = com-

plete–strand = both), and SNAP (2013-11-29) (parameters: -gff pasa1.hmm) were applied to predict the built-in *P. capsici* genome. For homology-based protein prediction, a set of protein sequences (anchored chromosome level) of LT1534[15], LT263[16], SD33[17], JS2[7], and T30-4[9] were downloaded and combined from the *Phytophthora* public database. Protein sequences were aligned to the genome using TblastN (v2.2.26; E-value ≤ 1e-5), and then the matching proteins were aligned to the homologous genome sequences for accurate spliced alignments with GeneWisev2.4.1[38] software which was used to predict gene structure contained in each protein region. For transcriptomic data, the RNA-seq reads of BYA5 were aligned to the assembled genome using Hisat2 v2.2.1[39]. Reference-based assembly of transcripts were generated using Trinity v2.8.5[40] (parameters:–normalize_reads–full_cleanup–min_glue 2–min_kmer_cov 2). The non-redundant reference gene set was generated by merging genes predicted by three methods with EvidenceModeler v1.1.1 using PASA (Program to Assemble Spliced Alignment) terminal exon support and including masked transposable elements as input into gene prediction[41]. Moreover, the non-coding RNAs in the BYA5 genome were predicted by Rfam/Infernal v1.1.5[42] (http://infernal.janelia.org/) and tRNAscan-SE v.1.4[43] (http://lowelab.ucsc.edu/tRNAscanSE/).

Oomycete effector prediction. For effector prediction, Effectors v3.0[44] and SignalP v6.0[45] were independently used to identify putative secreted proteins from the BYA5 proteome, and the overlapping effectors from the two predictions were obtained. The candidate effectorome was then scanned using TMHMM v2.0[46] to exclude those containing transmembrane helices, yielding the final set of predicted effectors.

Data Records

The raw genomic sequencing data, and RNA-seq data in this study have been deposited in the National Center for Biotechnology Information (NCBI) under BioProject number (PRJNA1248058)[47]. The raw genomic data are available under accession numbers SRR33296427[48], SRR33296426[49], and SRR33296420[50], respectively. The final assembled genome is deposited under National Center for Biotechnology Information (NCBI) under the accession number of GCA_053541045.1[51]. The raw genomic sequencing data, and RNA-seq data are also available at the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, under BioProject accession number CRA025046[52]. The final assembled genome is deposited under Genome Warehouse of National Genomics Data Center (https://ngdc.cncb.ac.cn/) under the accession number of GWHFSSF00000000.1[53]. The genome annotations including CDS and protein-coding regions files have been submitted to the online open access repository Figshare[54].

Technical Validation

Manual adjustment of misjoin and detection of potentially contaminated sequences. To obtain a nearly complete and error-free reference genome, we manually corrected the misassembly and removed redundant contigs using Hi-C reads alignment within JuiceBox visualization. Furthermore, all centromeres and most telomeres (32/36) (TTTAGGG) are captured via Centromics v0.3, quarTeT v1.2[55] and Tidk v0.2.0[56] (parameters: tidk explore -t 8 -m 5 -x 12 tidk search -s TTTAGGG).

Sequencing data assessment. The short read data were assessed by Merqury v1.3[25] and BWA v0.7.8[57]. The genomic short sequencing reads had 49.87% GC content. The Q20 and Q30 percentages were 97.47% and 93.44%, respectively. The Hi-C sequencing data had 50.30% GC content, and had quality scores of 97.37% (Q20) and 92.75% (Q30), respectively.

Evaluation of the genome assembly. The genome assembly quality was evaluated through the Benchmarking Universal Single-Copy Orthologs (BUSCO) tool with the "eukaryota_odb10" lineage as a reference dataset. The results showed that 94.1% of all 255 BUSCO markers were assembled, and a consensus quality value (QV) score of 42.68, all refecting the high accuracy and completeness of the genome assembly.

Data availability

The authors confirm that the data supporting the findings of this study are available in the National Genomics Data Center (NGDC), and National Center for Biotechnology Information (NCBI), and figshare, can be accessed via. https://ngdc.cncb.ac.cn/gwh/Assembly/92621/show (assembled genome in NGDC). https://ngdc.cncb.ac.cn/gsa/browse/CRA025046 (raw data in NGDC). https://identifiers.org/ncbi/insdc.gca:GCA_053541045.1 (assembled genome in NCBI). https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1248058 (raw data in NCBI). https://doi.org/10.6084/m9.figshare.28856261 (genome annotations archive in figshare).

Code availability

All commands and pipelines used in data processing were executed according to the manual and protocols of the corresponding bioinformatic software and described in the Methods section, along with the versions. No custom code was generated for these analyses.

Received: 28 April 2025; Accepted: 18 December 2025;
Published online: 05 January 2026

References

1. Lamour, K. H., Stam, R., Jupe, J. & Huitema, E. The oomycete broad-host-range pathogen *Phytophthora capsici*. *Molecular Plant Pathology* **13**, 329–337, https://doi.org/10.1111/j.1364-3703.2011.00754.x (2012).
2. Kamoun, S. *et al*. The Top 10 oomycete pathogens in molecular plant pathology. *Molecular Plant Pathology* **16**, 413–434, https://doi.org/10.1111/mpp.12190 (2015).

3. Kamoun, S. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annual Review Of Phytopathology* **44**, 41–60, https://doi.org/10.1146/annurev.phyto.44.070505.143436 (2006).
4. Białas, A. *et al*. Lessons in Effector and NLR Biology of Plant-Microbe Systems. *Molecular Plant-Microbe Interactions: MPMI* **31**, 34–45, https://doi.org/10.1094/MPMI-08-17-0196-FI (2018).
5. Anderson, R. G., Deb, D., Fedkenheuer, K. & McDowell, J. M. Recent Progress in RXLR Effector Research. *Molecular Plant-Microbe Interactions: MPMI* **28**, 1063–1072, https://doi.org/10.1094/MPMI-01-15-0022-CR (2015).
6. Birch, P. R. J., Rehmany, A. P., Pritchard, L., Kamoun, S. & Beynon, J. L. Trafficking arms: oomycete effectors enter host plant cells. *Trends in Microbiology* **14**, 8–11, https://doi.org/10.1016/j.tim.2005.11.007 (2006).
7. Zhang, Z. *et al*. Complete telomere-to-telomere genomes uncover virulence evolution conferred by chromosome fusion in oomycete plant pathogens. *Nature Communications* **15**, 4624, https://doi.org/10.1038/s41467-024-49061-y (2024).
8. Tsykun, T. *et al*. High-Quality Genome of the Tree Pathogen *Phytophthora plurivora*—A Novel Resource for Epidemiological Research. *PhytoFrontiersTM* **3**, 888–892, https://doi.org/10.1094/PHYTOFR-05-23-0065-A (2023).
9. Matson, M. E. H., Liang, Q., Lonardi, S. & Judelson, H. S. Karyotype variation, spontaneous genome rearrangements affecting chemical insensitivity, and expression level polymorphisms in the plant pathogen *Phytophthora infestans* revealed using its first chromosome-scale assembly. *PLOS Pathogens* **18**, e1010869, https://doi.org/10.1371/journal.ppat.1010869 (2022).
10. Cox, M. P. *et al*. Chromosome-level assembly of the *Phytophthora agathidicida* genome reveals adaptation in effector gene families. *Frontiers in Microbiology* **13**, 1038444, https://doi.org/10.3389/fmicb.2022.1038444 (2022).
11. Lamour, K. H. *et al*. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Molecular Plant-Microbe Interactions: MPMI* **25**, 1350–1360, https://doi.org/10.1094/MPMI-02-12-0028-R (2012).
12. Reyes-Tena, A. *et al*. Genome Sequence Data of Six Isolates of *Phytophthora capsici* from Mexico. *Molecular Plant-Microbe Interactions: MPMI* **32**, 1267–1269, https://doi.org/10.1094/MPMI-01-19-0014-A (2019).
13. Lee, J. H., Siddique, M. I., Kwon, J. K. & Kang, B.-C. Comparative Genomic Analysis Reveals Genetic Variation and Adaptive Evolution in the Pathogenicity-Related Genes of *Phytophthora capsici*. *Frontiers in Microbiology* **12**, 69413, https://doi.org/10.3389/fmicb.2021.694136 (2021).
14. Villanueva, O., Nguyen, H. D. T. & Ellouze, W. Comparative Genomic and Secretome Analysis of *Phytophthora capsici* Strains: Exploring Pathogenicity and Evolutionary Dynamics. *Agronomy* **14**, 2623, https://doi.org/10.3390/agronomy14112623 (2024).
15. Szadkowski, E. *et al*. *Phytophthora capsici* genome assembly for two isolates using long-read Oxford Nanopore Technology sequencing. *Microbiology Resource Announcements* **12**, e0019623, https://doi.org/10.1128/MRA.00196-23 (2023).
16. Cui, C., Herlihy, J. H., Bombarely, A., McDowell, J. M. & Haak, D. C. Draft Assembly of *Phytophthora capsici* from Long-Read Sequencing Uncovers Complexity. *Molecular Plant-Microbe Interactions: MPMI* **32**, 1559–1563, https://doi.org/10.1094/MPMI-04-19-0103-TA (2019).
17. Shi, J. *et al*. Improved Whole-Genome Sequence of *Phytophthora capsici* Generated by Long-Read Sequencing. *Molecular Plant-Microbe Interactions: MPMI* **34**, 866–869, https://doi.org/10.1094/MPMI-12-20-0356-A (2021).
18. A, J. *et al*. Pathogenomics Insights into *Phytophthora capsici* and *Phytophthora* tropicalis-Sibling Species Causing Black Pepper Foot Rot: Genomic Architecture. *Metabolic Pathways, and Effector Diversity*. *Gene* **947**, 149328, https://doi.org/10.1016/j.gene.2025.149328 (2025).
19. Wang, W. *et al*. Functional Analysis of the C-5 Sterol Desaturase PcErg3 in the Sterol Auxotrophic Oomycete Pathogen *Phytophthora capsici*. *Frontiers in Microbiology* **10**, 13, https://doi.org/10.3389/fmicb.2022.811132 (2022).
20. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170–175, https://doi.org/10.1038/s41592-020-01056-5 (2021).
21. Winkworth, R. *et al*. Comparative Analyses of Complete *Peronosporaceae* (Oomycota) Mitogenome Sequences-Insights into Structural Evolution and Phylogeny. *Genome Biology and Evolution* **14**, evac049, https://doi.org/10.1093/gbe/evac049 (2022).
22. Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature Plants* **5**, 833–845, https://doi.org/10.1038/s41477-019-0487-8 (2019).
23. Durand, N. C. *et al*. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**, 99–101, https://doi.org/10.1016/j.cels.2015.07.012 (2016).
24. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology* **17**, 239, https://doi.org/10.1186/s13059-016-1103-0 (2016).
25. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**, 245, https://doi.org/10.1186/s13059-020-02134-9 (2020).
26. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**, W435–439, https://doi.org/10.1093/nar/gkl200 (2006).
27. Wang, Y., Pruitt, R. N., Nürnberger, T. & Wang, Y. Evasion of plant immunity by microbial pathogens. *Nature Reviews Microbiology* **20**, 449–464, https://doi.org/10.1038/s41579-022-00710-3 (2022).
28. Wilson, R. A. & McDowell, J. M. Recent advances in understanding of fungal and oomycete effectors. *Current Opinion In Plant Biology* **68**, 102228, https://doi.org/10.1016/j.pbi.2022.102228 (2022).
29. Fernandez, J. The Phantom Menace: latest findings on effector biology in the rice blast fungus. *aBIOTECH* **4**, 140–154, https://doi.org/10.1007/s42994-023-00099-4 (2023).
30. Zheng, Z. *et al*. Atypical RXLR effectors are involved in *Phytophthora cactorum* pathogenesis. *aBIOTECH* **6**, 50–62, https://doi.org/10.1007/s42994-025-00198-4 (2025).
31. Miao, J., Chi, Y., Lin, D., Tyler, B. M. & Liu, X. Mutations in ORP1 Conferring Oxathiapiprolin Resistance Confirmed by Genome Editing using CRISPR/Cas9 in *Phytophthora capsici* and *P. sojae*. *Phytopathology* **108**, 1412–1419, https://doi.org/10.1094/PHYTO-01-18-0010-R (2018).
32. Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S. & Thompson, W. F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols* **1**, 2320–2325, https://doi.org/10.1038/nprot.2006.384 (2006).
33. Belton, J. M. *et al*. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods San Diego Calif* **58**, 268–276, https://doi.org/10.1016/j.ymeth.2012.05.001 (2012).
34. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011 (2011).
35. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890, https://doi.org/10.1093/bioinformatics/bty560 (2018).
36. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* **37**, 907–915, https://doi.org/10.1038/s41587-019-0201-4 (2019).
37. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *PNAS* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
38. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Research* **14**, 988–995, https://doi.org/10.1101/gr.1865504 (2004).
39. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* **11**, 1650–1667, https://doi.org/10.1038/nprot.2016.095 (2016).

40. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644–652, https://doi.org/10.1038/nbt.1883 (2011).
41. Haas, B. J. *et al*. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biology* **9**, R7, https://doi.org/10.1186/gb-2008-9-1-r7 (2008).
42. Griffiths-Jones, S. *et al*. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Research* **33**, D121–124, https://doi.org/10.1093/nar/gki081 (2005).
43. 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 Research* **49**, 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
44. Sperschneider, J. & Dodds, P. N. EffectorP 3.0: Prediction of Apoplastic and Cytoplasmic Effectors in Fungi and Oomycetes. *Molecular Plant-Microbe Interactions*. *MPMI* **35**, 146–156, https://doi.org/10.1094/mpmi-08-21-0201-r (2022).
45. Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340**, 783–795, https://doi.org/10.1016/j.jmb.2004.05.028 (2004).
46. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* **305**, 567–580, https://doi.org/10.1006/jmbi.2000.4315 (2001).
47. Wan, Y. This study aimed to obtain high quality genomic sequence of *Phytophthora capsica* isolate BYA5. *BioProject* https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1248058 (2025).
48. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33296427 (2025).
49. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33296426 (2025).
50. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR33296420 (2025).
51. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_053541045.1 (2025).
52. *NGDC Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA025046 (2025).
53. *NGDC Genome Sequence Archive* https://ngdc.cncb.ac.cn/gwh/Assembly/92621/show (2025).
54. Wan, Y. The annotated file for *Phytophthora capsici* isolate BYA5. *Figshare* https://doi.org/10.6084/m9.figshare.28856261 (2025).
55. Lin, Y. *et al*. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture Research* **10**, uhad127, https://doi.org/10.1093/hr/uhad127 (2023).
56. Brown, M. R., Manuel Gonzalez de La Rosa, P. & Blaxter, M. tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics* **41**, btaf049, https://doi.org/10.1093/bioinformatics/btaf049 (2025).
57. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).

Acknowledgements

This work was supported by Taishan Scholar Foundation of Shandong Province (tsqn202103162, tsqn202211093), Key R&D Program of Shandong Province, China (2024CXPT03), Yunnan Department of Science and Technology-Development and adoption of innovative 'green' management tools against insect pest and disease on tomato and pepper in Yunnan (No. 202502AQ370001), the Natural Science Foundation of Shandong Province (SYS202206), National Natural Science Foundation of China (32172387). We thank Dr. Xili Liu of Northwest Agriculture and Forestry University for providing *P. capsici* isolate BYA5. We thank Dr. Daolong Dou of Nanjing Agricultural University for providing the genomic data and GFF annotation files of *P. capsici* isolate LT263. This work was supported by the bioinformatics services from Novogene Bioinformatics Technology Co., Ltd. (Tianjin, China), particularly in data processing and genomic annotation.

Author contributions

Y.C. and Q.W. conceived and supervised the project. Y.W., F.Z., C.X. and X.L. collected the samples and strain preservation; M.Z. and Y.W. performed the genomic data analysis; S.G. submitted the genome assembly data to the NCBI database. Y.W. and F.Z. drafted the initial version of the manuscript. Q.W. and Y.C. finalized the manuscript. All authors read and prove the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-025-06501-8.

Correspondence and requests for materials should be addressed to Q.W. or Y.C.