

## TECHNICAL ADVANCE

# Draft Assembly of *Phytophthora capsici* from Long-Read Sequencing Uncovers Complexity

Chenming Cui, John H. Herlihy, Aureliano Bombarely, John M. McDowell, and David C. Haak<sup>†</sup>

School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA 24061, U.S.A.

Accepted 1 September 2019.

Resolving complex plant pathogen genomes is important for identifying the genomic shifts associated with rapid adaptation to selective agents such as hosts and fungicides, yet assembling these genomes remains challenging and expensive. *Phytophthora capsici* is an important, globally distributed plant pathogen that exhibits widespread fungicide resistance and a broad host range. As with other pathogenic oomycetes, *P. capsici* has a complex life history and a complex genome. Here, we leverage Oxford Nanopore Technologies and existing short-read resources to rapidly generate a low-cost, improved assembly. We generated 10 Gbp from a single MinION flow cell resulting in >1.25 million reads with an N<sub>50</sub> of 13 kb. The resulting assembly is 95.2 Mbp in 424 scaffolds with an N<sub>50</sub> length of 313 kb. This assembly is approximately 30 Mbp bigger than the current reference genome of 64 Mbp. We confirmed this larger genome size using flow cytometry, with an estimated size of 110 Mbp. BUSCO analysis identified 97.4% complete orthologs (19.2% duplicated). Evolutionary analysis supports a recent whole-genome duplication in this group. Our work provides a blueprint for rapidly integrating benchtop long-read sequencing with existing short-read data, to dramatically improve assembly quality and integrity of complex genomes and offer novel insights into pathogen genome function and evolution.

**Keywords:** genomics, metabolomics, miniasm, MinION, nanopore sequencing, oomycete, pathogen, proteomics

All data associated with this project have been deposited with the Sequence Read Archive (PRJNA557142). The assembly, coding sequence, and annotation files are publicly available at VTechData under doi: 10.7294/83VW-6Z31.

Current address of A. Bombarely: Department of Bioscience/Dipartimento di Bioscienze University of Milan/Università degli Studi di Milano (UNIMI), III piano / torre B Via Celoria, 26 Milano, 20133, Italy.

<sup>†</sup>Corresponding author: D. C. Haak; [dhaak@vt.edu](mailto:dhaak@vt.edu)

**Funding:** This work was supported by a grant to D. C. Haak, A. Bombarely, J. M. McDowell from the Fralin Life Science Institute at Virginia Tech.

\*The e-Xtra logo stands for “electronic extra” and indicates that four supplementary figures and one supplementary table are published online.

The author(s) declare no conflict of interest.

With one of the widest host ranges in the genus, *Phytophthora capsici* is an important, destructive pathogen in diverse cropping systems, including, pepper, tomato, and potato (Lamour et al. 2012b; Leonian 1922). To understand the mechanisms involved in host adaptation, a Sanger sequencing-based reference genome was generated in 2012, using inbred isolate LT1534 (Lamour et al. 2012a). Comparative genomics revealed that, like other *Phytophthora* spp., *P. capsici* has a diverse array of effector proteins that are associated with pathogenicity (Lamour et al. 2012a; Stam et al. 2013). Tyler et al. (2006) and others have demonstrated that, in *Phytophthora* spp., virulence genes are often located in gene-poor regions interspersed within repetitive regions (Raffaele and Kamoun 2012). Resolving genomes at this level is important for identifying key genomic regions associated with pathogenicity traits. However, these complex regions of repeats are computationally challenging for using high-throughput short-read (150 to 300 bp) data because the reads often do not span the repetitive region, therein prohibiting accurate assembly.

Recent advances in sequencing technologies are improving our ability to generate long reads (3 to 15 kbp) that allow us to overcome these complexities by spanning more of the repetitive region, allowing accurate assembly and uncovering previously hidden genomic information. For example, using the Pacific Biosciences Sequel platform, Yang et al. (2018) generated a full-length assembly for *P. cactorum* that spanned 121.5 Mb and consisted of approximately 46% repetitive sequence. Novel findings from this sequencing project include the identification of a recent whole-genome duplication (WGD) and subsequent gene loss in this lineage, as well as expanded gene families associated with pathogenicity, relative to the more host-specific *P. sojae* (Yang et al. 2018). Similarly, Malar C et al. (2019) generated a haplotype-phased assembly of *P. ramorum* from PacBio long reads identifying an overall repeat increase from 29% (prior assembly) to 48% and described newly identified RXLR effector genes. The obvious downside of long-read sequencing is both the high per-base cost to generate the reads (depending on sequencing cell output) and the time and technical expertise needed for library preparation (Fletcher et al. 2019).

The Oxford Nanopore Technologies (ONT) MinION platform is a so-called third-generation, long-read platform that is small and aimed at lab-based users and requires minimal technical expertise. Importantly, sequencing flow cell output can be improved by the end user through adjustments to extraction and preparation protocols. Finally, the time from sample preparation to data acquisition is greatly diminished and eliminates the need for sample shipment to a sequencing facility. Here, we report the use of ONT MinION sequencing technology and a streamlined bioinformatics pipeline that includes minimap2 and miniasm

(Li 2016) to develop an improved reference genome for *P. capsici* in only 9 days. This cost-effective, improved assembly revealed novel gene-space and genomic complexity and enabled a substantial revision of *P. capsici* genome size.

# RESULTS

## Genome sequencing and assembly.

The isolate used in this study, LT263, was originally isolated from infected pumpkin in Tennessee in 2004. Sequencing was completed on the ONT MinION platform. A single 1D R9.4 flow cell generated 1,258,480 reads (approximately 10 Gb) at 70× with an N<sub>50</sub> (number of contigs comprising the 50th percentile of the data) read length of 11,507 bp; the longest read was 99,577 bp and the mean read length was 7,114 bp.

Base-called nanopore raw reads were assembled into 95.2 Mb using a custom bash pipeline that included overlapping via minimap2 (Li 2018), overlap consensus de novo assembly via miniasm (Li 2016), haplotype collapsing via purge\_haplotigs (Roach et al. 2018), and successive sequence polishing via Racon (Vaser et al. 2017). Because miniasm concatenates overlaps produced by minimap2 (or other overlappers), raw assemblies are contiguous but retain the error rate of the input reads. Thus, the *P. capsici* raw assembly was 94.4 Mb contained in 603 contigs (Table 1), with an N<sub>50</sub> length of 194.4 kb; however, it retained an approximately 14.4% error rate and, therefore, captured only 114 of 234 (48.7%) complete BUSCOs (Table 2) (alveolata\_stramenophiles\_ensembl database implemented in BUSCO v 3.0) (Simão et al. 2015).

Polishing the assembly with the uncorrected nanopore long reads using Racon v1.3.1 (Vaser et al. 2017) improved the BUSCO score to 145 of 234 (62.0%) (Table 2). Completeness was improved substantially to 228 of 234 (97.4%) complete BUSCOs (Table 2) after polishing with available short-read data from LT1534, a line derived from a cross between LT263 and OP97 which was then successively backcrossed to LT263 two times (Lamour et al. 2012a). BUSCO analysis on this final dataset revealed a duplication rate of 19.2% (Table 2). This increased duplication rate is in contrast with the published LT1534 assembly (Lamour et al. 2012a), with a single-copy BUSCO score of 91.0% and duplicated COGs at 0.00% (Table 2). The polished assembly was scaffolded using SSPACE-Long (Boetzer and Pirovano 2014), resulting in 424 scaffolds with an L<sub>50</sub> (average length of the contigs in the 50th percentile of the data) of 313 kbp, and 99% of scaffolds >10 kb (Table 1). Assembly, error correction, and scaffolding took just 72 h using 20 threads and 100 Gb of RAM on a 48-core server with 1 Tb of RAM available. In contrast, assembling corrected reads with Canu v1.7 (Koren et al. 2017) was completed in 132 h, and raw assembly with Canu v1.7 using internal read correction exceeded 30 days.

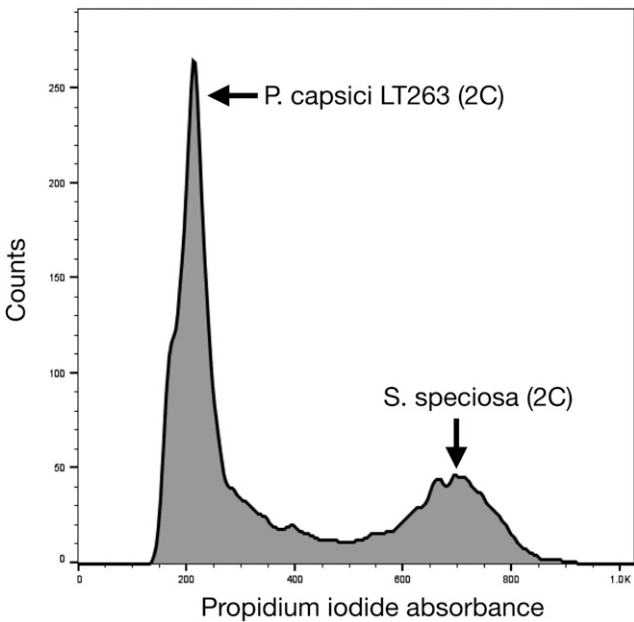
## Confirmation of *P. capsici* LT263 genome size.

The 95.2-Mb size of the assembly was surprising, because it is larger than the previous estimate of 64 Mb derived from *P. capsici* LT1534 (Lamour et al. 2012a). Flow cytometry was conducted to confirm the genome size of this isolate. Using *Sinningia speciosa* (1C size approximately 392 Mb) (Zaitlin and Pierce 2010) as an internal standard, the haploid nuclear content of this isolate was estimated at 110 Mb (Fig. 1).

## Genome sequence analysis and comparative genomics.

A combination of homology and ab initio methods implemented in Maker (Yandell and Ence 2012) predicted 19,391 protein coding regions from 424 scaffolds (Table 3). This corresponds well with the previous estimate of 19,805 (Lamour et al. 2012a), indicating that this assembly captured 99.98% of the gene space. This results in a gene density of 204 genes/Mb, which is within the range of other members of the genus (Yang et al. 2018). Thus, genomic architecture was not substantially altered, with strong synteny detected between the current assembly, reference genome, and other *Phytophthora* spp. genomes (Fig. 2), similar to other comparative studies (Lamour et al. 2012a).

A *P. capsici* paranome K<sub>s</sub> analysis (Fig. 3) supports the long-standing hypothesis of an ancestral whole-genome duplication



**Fig. 1.** Flow cytometry plot confirming the larger assembly size for *Phytophthora capsici* LT263. The first peak is *P. capsici* (approximately 110 Mb) and the second peak is an internal standard from *Sinningia speciosa* (approximately 392 Mb).

**Table 1.** Genome assembly data for the *Phytophthora capsici* genome<sup>a</sup>

Assembly	Total assembly (Mbp)	Number of contigs	Longest contig (Mbp)	N <sub>50</sub> of contigs	L <sub>50</sub> of contigs (kbp)
Scaffolds	95.18	424	1.44	95	313.00
Contigs	94.35	603	1.26	152	194.36

<sup>a</sup> N<sub>50</sub> = the number of contigs in the 50th percentile of the data; L<sub>50</sub> = the average length of the contigs in the 50th percentile of the data.

**Table 2.** Genome completeness summary data for the *Phytophthora capsici* genome

Completeness	Complete BUSCOs (%)	Single copy (%)	Duplicated (%)	Fragmented (%)	Missing (%)
Raw assembly	48.70	42.30	6.40	1.70	49.60
Long read polish	62.00	50.00	12.00	1.70	36.30
Short read polish	97.40	78.20	19.20	0.40	2.20
LT1534 assembly	91.00	91.00	0.00	2.10	6.90

leading to the clade containing *P. ramorum*, *P. infestans*, *P. sojae*, and *P. capsici* (Martens and Van de Peer 2010). A  $K_s$  plot with kernel density estimation reveals peaks consistent with recent small duplication events as well as ancestral larger duplication events (Fig. 3A). Bayesian mixture models with components from 1 to 8 were fitted to the  $K_s$  distribution (Supplementary Fig. S3). The best-fit model was selected based on Bayes factors and component weighting, resulting in a model with five log-normal distributions (Fig. 3B and D). The peak at  $K_s$  0.1 to 0.5 is consistent with the speciation event leading to the *P. capsici* lineage. A second peak at  $K_s$  1.7 to 2.0 is consistent with an ancestral large duplication event. These peaks are also supported in *P. capsici*–*P. sojae*  $K_s$  plots (Supplementary Fig. S4). Though these values of  $K_s$  are close to saturation, they are consistent with other studies supporting a WGD in this clade (Martens and Van de Peer 2010; Yang et al. 2018).

# DISCUSSION

A long-standing goal in genomics has been the development of sequencing and assembly approaches that allow the development of contiguous genome sequences in a reasonably short timeframe. We leveraged ONT MinION generated long-read and available short-read sequencing data to assemble the complex genome from *P. capsici*. In addition, using recently introduced algorithms, the assembly was completed in one-sixth of the time required for standard approaches. Importantly, we found that the genome was larger than previous estimates and represented an increase in repeat content.

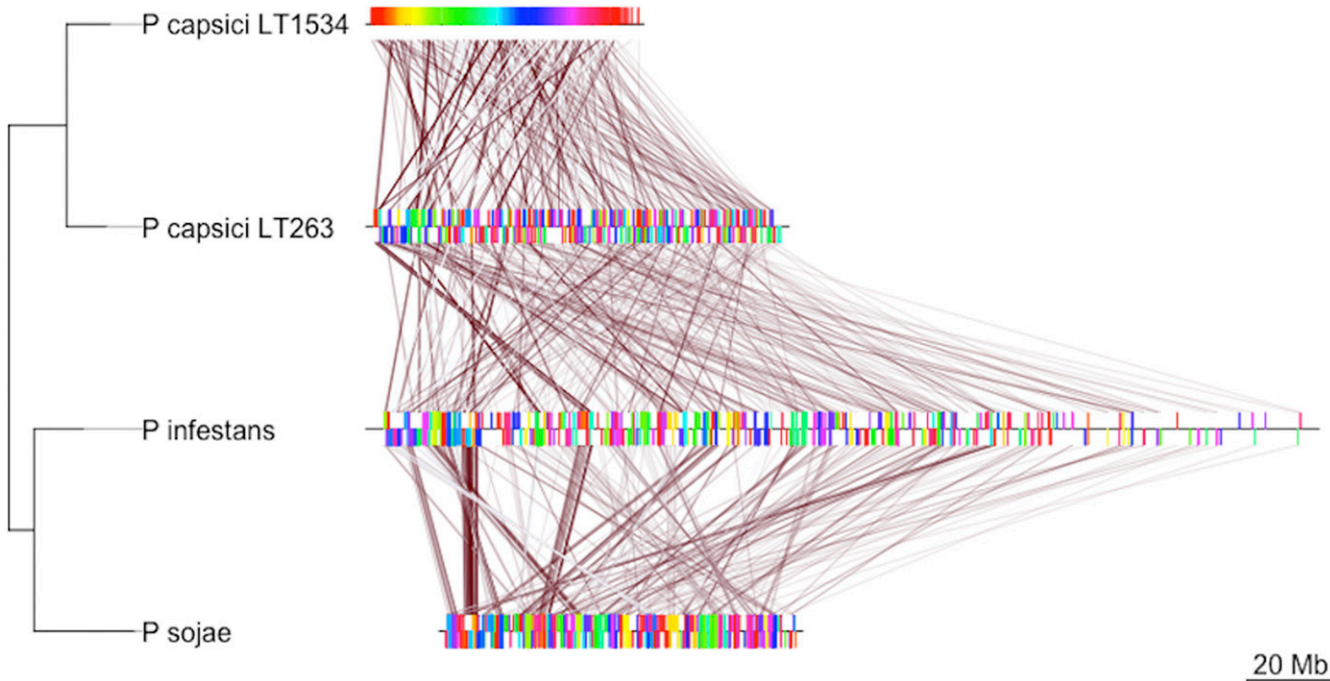
**Table 3.** Coding regions from 424 scaffolds

Parameters	Totals
GC content of the genome (%)	50.95
Number of predicted protein-coding genes	19,391
Number of protein-coding mRNAs	33,549

Although prior estimates of the genome size were much smaller, our assembly size was confirmed by a flow cytometry-estimated size of 110 Mb. Previous estimates were based on assemblies that, in part, used short-read data, where repetitive regions are often collapsed (Treangen and Salzberg 2011). Conversely, generating overlaps from high-error long-read data can lead to false expansion from partial overlaps at repeats (Chu et al. 2017). We anticipate that the repeat-rich regions in the new assembly will enable identification of additional genes with functions in plant host interactions, as shown recently for *P. ramorum* (Malar C et al. 2019).

Annotation of our assembly using available RNAseq data (Lamour et al. 2012a) resulted in the capture of 99.9% of the previously identified genes in this species. In addition, we report a substantial increase in the number of duplicated, conserved orthologous genes detected. These differences, however, were similar to long-read genome assemblies for other closely related members of the genus (Yang et al. 2018). The increase in repeats across the genome is most likely the result of a prior WGD, which is supported in data from other species (*P. ramorum*, *P. cactorum*, and *P. sojae*) (Malar C et al. 2019; Martens and Van de Peer 2010; Yang et al. 2018). For our assembly, this is supported by evidence from *P. capsici* paralog and *P. capsici*–*P. sojae*  $K_s$  plots, wherein we find  $K_s$  peaks that are consistent with recent duplication events and the retention of genes from an older large duplication event as well. Yang et al. (2018) found that this WGD lead to an expansion of gene families in the clade that contains *P. cactorum* and *P. sojae*. Further comparative genomic analysis within this clade will identify the timing of this WGD and its impact on gene family diversification leading to adaptation.

Resolving the complex genomes of plant pathogens is an important step toward understanding the mechanisms through which they adapt to host plants. We have coupled a lab-based sequencing approach and efficient assembly algorithms to generate a de novo assembly for *P. capsici* that captured previously



**Fig. 2.** Syntenic plot between *Phytophthora infestans*, *P. sojae*, *P. capsici* LT263, and *P. capsici* LT1534 with a guide tree based on nucleotide distance. Blocks were generated from progressive Mauve whole-genome alignments and filtered to retain blocks above 2 kbp. Colors indicate syntenic blocks ordered by *P. capsici* LT1534, and regions connected by lines are considered shared blocks between the genomes. The extended length of the horizontal line in the LT263 assembly represents the expanded genome.

undescribed complexity. An important part of this assembly was the availability of public data for genome polishing and annotation. We anticipate that costs associated with hybrid sequencing approaches such as those presented here will continue to decrease as an increasing number of plant pathogen sequencing projects are completed and those data are added to repositories. In total, our sequencing effort was just 9 days from DNA extraction to polished assembly, pushing us closer to “real-time” whole-genome sequencing of plant pathogens in the field.

## MATERIALS AND METHODS

### Strain selection and cultivation conditions.

*P. capsici* isolate LT263 is used globally for its virulence on a wide range of hosts, and its sexual and asexual fecundity is amenable for genetic studies. The isolate was obtained from Kurt Lamour, and maintained on 10% V8 agar plates at 26°C in the dark. For DNA extraction, flasks with 10% V8 liquid media were inoculated and grown in a shaker incubator at 26°C in the dark for 7 days. Hyphae were collected from the flasks in 50-ml centrifuge tubes, immediately frozen in liquid nitrogen, then stored at –80°C until use.

### DNA extraction and nanopore sequencing.

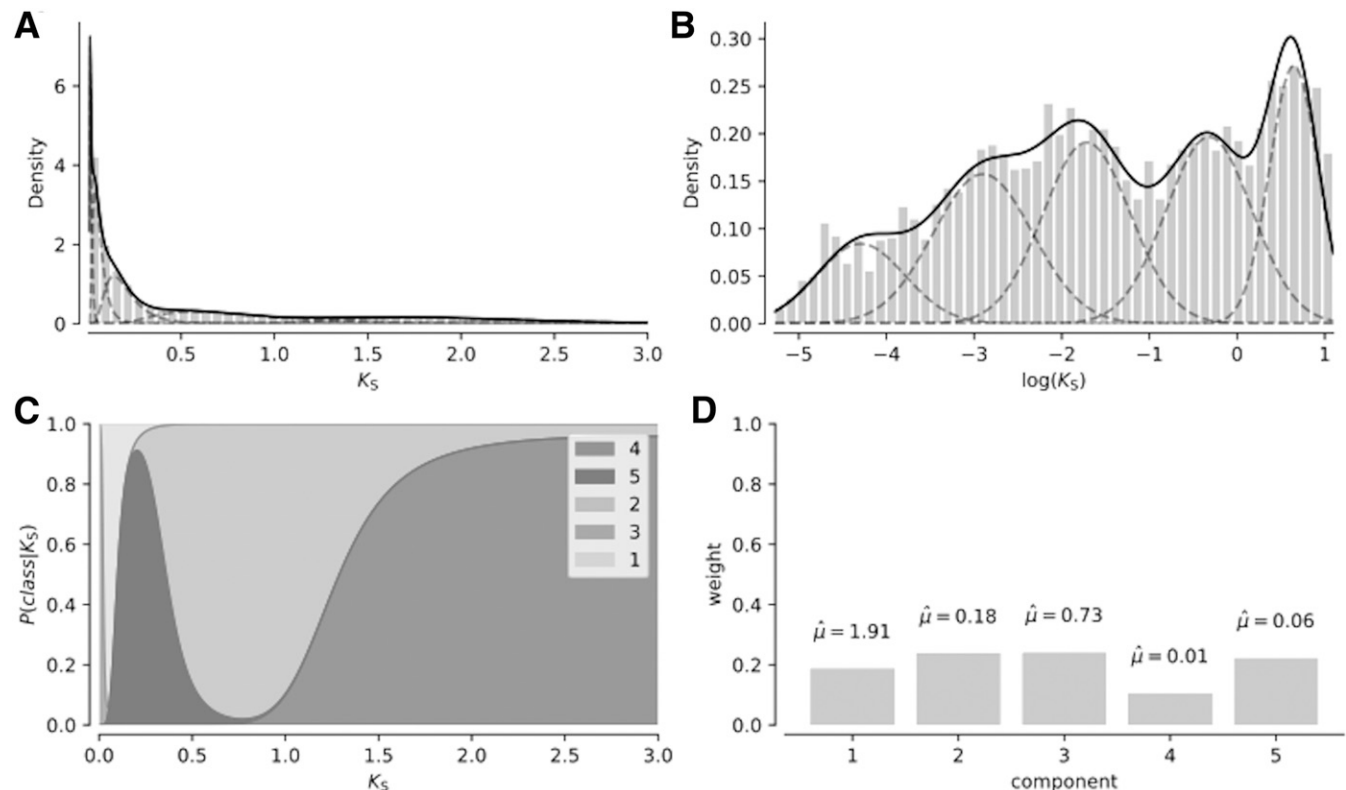
High-quality *P. capsici* DNA was isolated using a modified DNeasy Plant Mini Kit, according to the manufacturer’s instructions (Qiagen). The genomic DNA (gDNA) was sequenced using the MinION platform. Sequencing was preceded by library preparation from 1.5 µg of gDNA using a 1D Genomic DNA sequencing kit SQK-LSK108 from ONT. DNA fragmentation was not performed in order to retain longer fragments for sequencing. The extracted DNA was repaired using the FFPE Repair Mix (New England Biolabs), followed by end repair and dA-tailing using the NEBNext End Prep Module

(New England Biolabs). Then, the adapter was ligated to the cleaned DNA using Blunt/TA Ligase Master Mix (New England Biolabs). All bead washing steps were completed using AMPure XP beads (Beckman Coulter). The prepared libraries were sequenced using an R9.4 flow cell on the MinION device. Sequencing was performed using the 48 run time protocol of MinKNOW2.2 software. In total, 1.258 million reads, which translated to 10.06 Gb, were generated from a single flow cell.

### Genome assembly and assessment.

Base calling was performed using Albacore v2.3.1 (ONT) and subsequent raw FAST5 files were converted into a single combined FASTQ file. Read lengths that were <1,000 bp were filtered out prior to genome assembly. Raw reads were assembled via a custom bash pipeline that used minimap2 v2.12-r829-dirty (Li 2018) with settings -x ava-ont to generate overlaps, miniasm v0.3-r179 (Li 2016) to assemble overlaps, purge\_haplotigs (Roach et al. 2018) to resolve haplotypes, and two rounds of racon v1.3.1 polishing. The first round of polishing used the raw reads and a second round used publicly available Illumina sequence data. Because there are no available short-read data available for pure LT263, we used reads from LT1534 (PRJNA386483), which is a hybrid between LT263 and OP97 which was then backcrossed to LT263 twice. The polished assembly was scaffolded using SSPACE-Long v1-1 (Boetzer and Pirovano 2014) and the scaffolds were annotated using Maker v2.31.8 (Yandell and Ence 2012). For comparison, raw reads were also assembled using Canu v1.7 (Koren et al. 2017).

Genome scaffolds were assessed for integrity and length using Quast v4.6.3 (Gurevich et al. 2013) and the quality of gene space Δcapture was assessed using BUSCO v3.0 (Simão et al. 2015). Assembly, polishing, and scaffolding of version pcapsici\_VT1.1 were completed on a Linux server running Ubuntu v18.04 with 96 cores and 1 TB of RAM available.



**Fig. 3.** Parame synonymous substitution ( $K_s$ ) plots indicating the presence of duplication in this lineage. **A**, Kernel density estimation of peaks in the  $K_s$  distribution. **B**, Results of five log normal distributions in the  $K_s$  distribution from Bayesian gaussian mixture models (bgmm). **C**, Assignment probabilities for each of the five detected components from the bgmm analysis. **D**, Component weights from the bgmm analysis.

Evolutionary analyses were completed using Mauve snapshot\_2015\_2\_25 (1) (Darling et al. 2004) to generate whole-genome alignments and identify blocks locally colinear blocks. Mauve-generated backbone and guide tree files were used to generate synteny plots in genoPlotR v0.8.9 (Guy et al. 2010). The backbone file was filtered to remove blocks smaller than 2 kbp and set *P. capsici* LT1534 as the reference. Parame analysis of *P. capsici* LT263 was completed using wgd wf1 v1.0 (Zwaenepoel and Van de Peer 2019). Model comparisons were conducted using wgd mix bgmms with up to 10 components, with 10 k-means initializations across 10,000 iterations. Model selection was twofold, first based on Bayesian information criterion (BIC) (lowest  $\Delta$ BIC) and then selection by component weights.

### Genome size estimation.

We modified the protocol of Galbraith et al. (1983) for use with *P. capsici* in culture (Q. Zhang, M. Makris, J. H. Herlihy, and D. C. Haak, unpublished data). In short, *P. capsici* LT263 was maintained on 10% clarified V8 plates (1.5% agar) with  $\beta$ -sitosterol at 30 mg/liter, then transferred to the same liquid medium in a shaker incubator (28°C) and kept in the dark for 7 days. Approximately 0.8 to 1.2 g (wet weight) of hyphae was combined with 0.5 g (wet weight) of fresh leaves of *S. speciosa* (1C = 392 Mb) (Zaitlin and Pierce 2010) and co-chopped. After chopping, samples were filtered by columns and transferred to chilled mortar and ground for 15 s. After grinding, 1 ml of de Laat's buffer (de Laat and Blaas 1984) containing cell constituents was added and the resulting suspension was successively passed through a 50- $\mu$ m filter and a 10- $\mu$ m filter. A 1:1 (vol/vol) amount of staining solution containing propidium iodide (50  $\mu$ g/ $\mu$ l), RNase (50  $\mu$ g/ml), and  $\beta$ -mercaptoethanol (1.1  $\mu$ l/ml) was added. Samples were gently mixed and incubated in the dark at room temperature for 20 min. Samples were then kept in the dark at 4°C until processed in the Flow Cytometry Resource Laboratory at Virginia Tech. Relative fluorescence was measured with the FL2 detector, and DNA content was quantified with FL2-area (integrated fluorescence) and displayed on histograms (Baldwin and Husband 2013).

### ACKNOWLEDGMENTS

We thank Q. Zhang for her work developing a Flow Cytometry protocol for *P. capsica*, M. Makris and the Flow Cytometry Resource Lab in the Virginia Tech College of Veterinary Medicine for conducting the flow analyses, K. Lamour for providing *P. capsici* isolate LT263, and the *Phytophthora* community for providing public access to genomic datasets.

### AUTHOR-RECOMMENDED INTERNET RESOURCE

VTechData: <https://data.lib.vt.edu>

### LITERATURE CITED

Baldwin, S. J., and Husband, B. C. 2013. The association between polyploidy and clonal reproduction in diploid and tetraploid *Chamerion angustifolium*. *Mol. Ecol.* 22:1806-1819.

Boetzer, M., and Pirovano, W. 2014. SSPACE-LongRead: Scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinf.* 15:211.

Chu, J., Mohamadi, H., Warren, R. L., Yang, C., and Biról, I. 2017. Innovations and challenges in detecting long read overlaps: An evaluation of the state-of-the-art. *Bioinformatics* 33:1261-1270.

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14:1394-1403.

de Laat, A. M., and Blaas, J. 1984. Flow-cytometric characterization and sorting of plant chromosomes. *Theor. Appl. Genet.* 67:463-467.

Fletcher, K., Gil, J., Bertier, L. D., Kenefick, A., Wood, K. J., Zhang, L., Reyes-Chin-Wo, S., Cavanaugh, K., Tsuchida, C., Wong, J., and

Michellmore, R. 2019. Genomic signatures of somatic hybrid vigor due to heterokaryosis in the oomycete pathogen, *Bremia lactucae*. *bioRxiv* 516526.

Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, N. M., Sharma, D. P., and Firoozabady, E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220:1049-1051.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29:1072-1075.

Guy, L., Roat Kultima, J., and Andersson, S. G. E. 2010. genoPlotR: Comparative gene and genome visualization in R. *Bioinformatics* 26:2334-2335.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27:722-736.

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., Bharti, A. K., Donahoo, R. S., Finley, S., Huitema, E., Hulvey, J., Platt, D., Salamov, A., Savidor, A., Sharma, R., Stam, R., Storey, D., Thines, M., Win, J., Haas, B. J., Dinwiddie, D. L., Jenkins, J., Knight, J. R., Affourtit, J. P., Han, C. S., Chertkov, O., Lindquist, E. A., Detter, C., Grigoriev, I. V., Kamoun, S., and Kingsmore, S. F. 2012a. Genome sequencing and mapping reveal loss of heterozygosity as a mechanism for rapid adaptation in the vegetable pathogen *Phytophthora capsici*. *Mol. Plant-Microbe Interact.* 25:1350-1360.

Lamour, K. H., Stam, R., Jupe, J., and Huitema, E. 2012b. The oomycete broad-host-range pathogen *Phytophthora capsici*. *Mol. Plant Pathol.* 13:329-337.

Leonian, L. H. 1922. Stem and fruit blight of peppers caused by *Phytophthora capsici* sp. *Phytopathology* 12:401-408.

Li, H. 2016. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32:2103-2110.

Li, H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094-3100.

Malar C, M., Yuzon, J. D., Das, S., Das, A., Panda, A., Ghosh, S., Tyler, B. M., Kasuga, T., and Tripathy, S. 2019. Haplotype-phased genome assembly of virulent *Phytophthora ramorum* isolate ND886 facilitated by long-read sequencing reveals effector polymorphisms and copy number variation. *Mol. Plant-Microbe Interact.* 32:1047-1060.

Martens, C., and Van de Peer, Y. 2010. The hidden duplication past of the plant pathogen *Phytophthora* and its consequences for infection. *BMC Genomics* 11:353.

Raffaele, S., and Kamoun, S. 2012. Genome evolution in filamentous plant pathogens: Why bigger can be better. *Nat. Rev. Microbiol.* 10:417-430.

Roach, M. J., Schmidt, S. A., and Borneman, A. R. 2018. Purge haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinf.* 19:460.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.

Stam, R., Jupe, J., Howden, A. J. M., Morris, J. A., Boevink, P. C., Hedley, P. E., and Huitema, E. 2013. Identification and characterisation CRN effectors in *Phytophthora capsici* shows modularity and functional diversity. *PLoS One* 25:e59517.

Treangen, T. J., and Salzberg, S. L. 2011. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nat. Rev. Genet.* 13:36-46.

Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., and Chapman, J. 2006. *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313:1261-1266.

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27:737-746.

Yandell, M., and Ence, D. 2012. A beginner's guide to eukaryotic genome annotation. *Nat. Rev. Genet.* 13:329-342.

Yang, M., Duan, S., Mei, X., Huang, H., Chen, W., Liu, Y., Guo, C., Yang, T., Wei, W., Liu, X., He, X., Dong, Y., and Zhu, S. 2018. The *Phytophthora cactorum* genome provides insights into the adaptation to host defense compounds and fungicides. *Sci. Rep.* 8:6534.

Zaitlin, D., and Pierce, A. J. 2010. Nuclear DNA content in *Sinningia* (Gesneriaceae); intraspecific genome size variation and genome characterization in *S. speciosa*. *Genome* 53:1066-1082.

Zwaenepoel, A., and Van de Peer, Y. 2019. wgd—Simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35:2153-2155.