

High-Quality Reference Genome Sequence for the Oomycete Vegetable Pathogen *Phytophthora capsici* Strain LT1534

 Jason E. Stajich,^{a,b} Andrea L. Vu,^{a,b*}  Howard S. Judelson,^{a,b} Gregory M. Vogel,^c Michael A. Gore,^c Maryn O. Carlson,^{c*} Nicholas Devitt,^d Jennifer Jacobi,^d Joann Mudge,^d Kurt H. Lamour,^e Christine D. Smart^c

^aDepartment of Microbiology and Plant Pathology, University of California, Riverside, California, USA

^bInstitute for Integrative Genome Biology, University of California, Riverside, California, USA

^cSchool of Integrative Plant Science, Cornell University, Geneva, New York, USA

^dNational Center for Genome Resources, Santa Fe, New Mexico, USA

^eDepartment of Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee, USA

ABSTRACT The oomycete *Phytophthora capsici* is a destructive pathogen of a wide range of vegetable hosts, especially peppers and cucurbits. A 94.17-Mb genome assembly was constructed using PacBio and Illumina data and annotated with support from transcriptome sequencing (RNA-Seq) reads.

Phytophthora capsici is a highly destructive pathogen of vegetables worldwide, especially those in the Solanaceae and Cucurbitaceae families (Fig. 1). Efforts to understand the role of genes in pathogenesis and host range (1–8), population structure (9, 10), genomic variation (11), and development (12) benefit from an accurate genome assembly. *P. capsici* belongs to the eukaryotic phylum Oomycota, many members of which have repeat-rich genomes that are difficult to assemble with short reads. Strain LT1534, A2 mating type, is an inbred strain derived from crossing isolates from infected cucurbits in Michigan (cucumber) and Tennessee (pumpkin) (11). An earlier assembly for LT1534, generated with Roche 454 and Sanger sequencing (11), was 64 Mb in 917 scaffolds (N_{50} , 706 kb).

A single oospore-derived isolate (LT1534) was maintained axenically on PARP (25 ppm pimarin, 100 ppm ampicillin, 25 ppm rifampicin, and 100 ppm pentachloronitrobenzene) agar plates, and a small weft of mycelium was transferred to clarified V8 juice broth amended with PARP and grown at 22 to 23°C for 7 days prior to harvesting, freeze-drying, disruption, and extraction of genomic DNA (13) using a GeneJet genomic DNA purification kit (Thermo Fisher). Libraries were constructed with a TruSeq DNA kit and sequenced on an Illumina HiSeq X Ten system by Novogene (Shenzhen, China), yielding 34.9 million, 2×150 -bp read pairs (10.48 Gbp). PacBio libraries were constructed by the National Center for Genome Resources (Santa Fe, NM) following the 20-kb protocol (14). One single-molecule real-time (SMRT) cell was sequenced on a PacBio RS II system using P6 polymerase and C4 chemistry, yielding 964,374 subreads ($n = 964,374$; N_{50} , 12.5 kb; maximum, 47.5 kb; total, 8.15 Gbp) as processed by the SMRT pipeline v2.3.0.1. PacBio reads were corrected with Illumina by LorDEC v0.9 (15) using the default parameters.

To provide gene model support, mRNA was extracted from 2.5-, 3-, and 4.5-day V8 agar cultures grown at 23°C under 12-h light/dark conditions using a Spectrum plant total RNA kit (Sigma). Strand-specific transcriptome sequencing (RNA-Seq) libraries were constructed and sequenced by Cofactor Genomics (St. Louis, MO) using oligo(dT) priming on an Illumina NextSeq 500 instrument to obtain ~25 million single-end reads per library.

Read trimming, correction, and genome assembly were performed with MaSuRCA v3.3.1 (16, 17) with Illumina and LorDEC-corrected PacBio reads (LHE_COVERAGE=25

Citation Stajich JE, Vu AL, Judelson HS, Vogel GM, Gore MA, Carlson MO, Devitt N, Jacobi J, Mudge J, Lamour KH, Smart CD. 2021. High-quality reference genome sequence for the oomycete vegetable pathogen *Phytophthora capsici* strain LT1534. Microbiol Resour Announc 10:e00295-21. <https://doi.org/10.1128/MRA.00295-21>.

Editor Antonis Rokas, Vanderbilt University

Copyright © 2021 Stajich et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Jason E. Stajich, jason.stajich@ucr.edu.

* Present address: Andrea L. Vu, Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, North Carolina, USA; Maryn O. Carlson, Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, Illinois, USA.

Received 22 March 2021

Accepted 7 May 2021

Published 27 May 2021

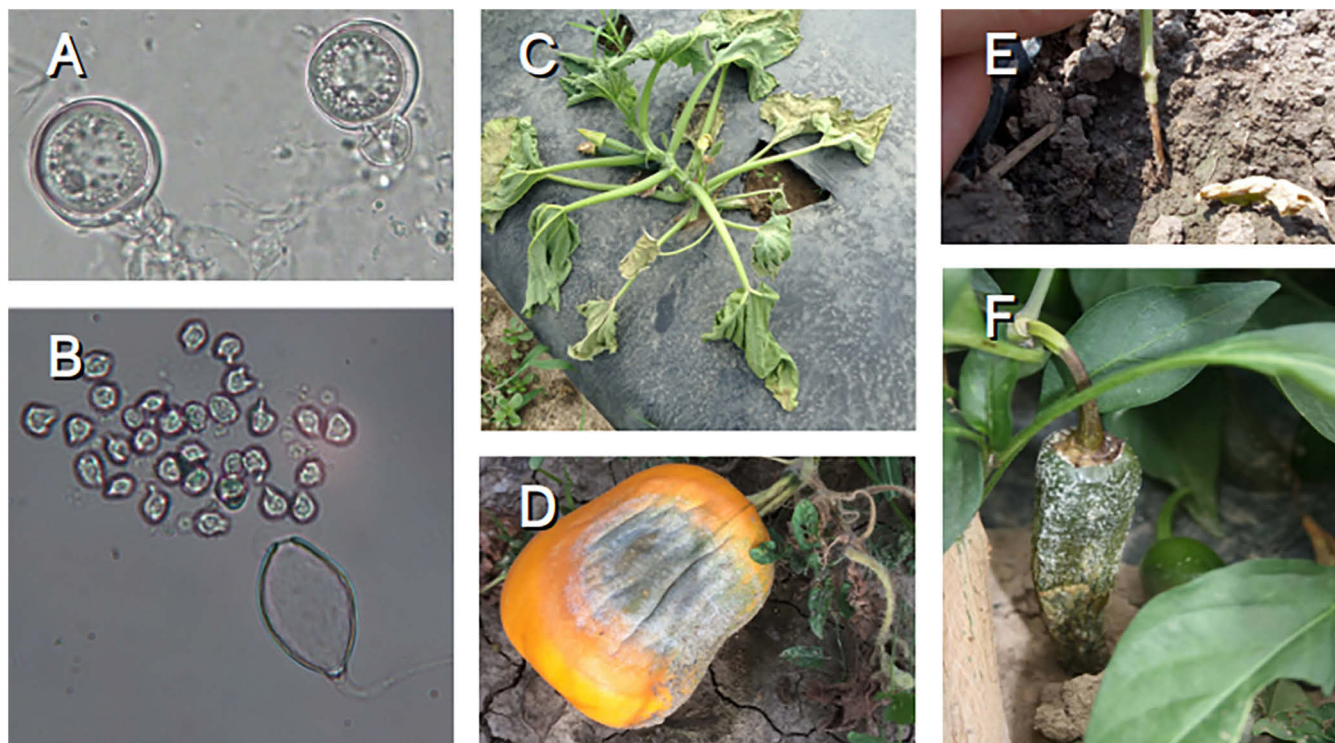


FIG 1 Micrographs of *P. capsici* reproductive structures and examples of disease symptoms. (A) Oospores, the sexual overwintering structures, with amphigynous antheridia in view. (B) A sporangium releasing motile zoospores. The sporangiophore can be seen attached to the base of the sporangium. (C) Crown rot on zucchini causing total plant wilt and death. (D) Fruit rot on pumpkin with visible white sporangia. (E) Stem of young pepper plant showing brown girdling due to crown rot. (F) Fruit rot on jalapeno pepper. Symptoms started at the tip of the fruit, which is touching the soil, and moved up the fruit. Typical white symptoms are sporangia of *P. capsici* with each sporangium on a sporangiophore. There would be thousands of sporangia per infected fruit similar to the one shown here.

cgwErrorRate=0.15). The assembly was screened for adaptors and contamination using AATF v0.2.3 (18). Assembly polishing with masurca-polish.sh corrected 1,025 substitutions and 2,643 indels, resulting in 99.9961% computed consensus quality. The 94.17-Mb assembly is in 782 scaffolds (N_{50} , 485 kb; L_{50} , 44; average GC content, 52.3%), and the longest scaffold covers 4.61 Mb. Completeness was assessed with BUSCO v3.0.2 (19, 20) using protists_ensembl v9, resulting in the identification of 211 (98.1%) out of 215 genes; 175 were single-copy complete, 36 were duplicated, and 2 were fragmented.

Genome annotation was performed with Funannotate v1.8.1 (21) using default parameters, which implemented the following steps. Prediction training and annotation were supported by RNA-Seq reads aligned to the genome with HISAT2 v2.2.1 (22) and reference-guided transcript assembly in Trinity v2.11.0 (maxintron=4kb) (23, 24) and PASA v2.4.1 (25). The best gene models were used to train and run SNAP v2013_11_29 (26) and AUGUSTUS v3.3.3 (27). Additional *ab initio* models were predicted using GeneMark v4.59 (28), GlimmerHMM v3.0.4 (29), and CodingQuarry v2.0 (30). Evidence for exons was generated by DIAMOND v2.0.4 (31) and Exonerate v2.4.0 (32) alignments of SwissprotDB (33) proteins. Consensus gene models were produced with EVIDENCEModeler v1.1.1 (25) using Funannotate default evidence weights. Untranslated regions and alternatively spliced isoforms were predicted using PASA from RNA-Seq. The putative protein function was assigned by sequence similarity to the InterProScan v5.45-80.0 (34), eggNOG v1.0.3 (35), dbCAN2 v9.0 (36), and MEROPS v12.0 (37) databases. The genome has 23,373 predicted protein-coding genes, 133 of which had at least one putatively alternatively spliced isoform.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank as accession number [JADEVP000000000](#). The version described in this paper is version [JADEVP010000000](#). The PacBio (SRA number [SRR13176613](#)) and Illumina (SRA

number [SRR13176614](#)) genomic sequencing reads are associated with BioProject [PRJNA481983](#). The RNA-Seq reads are associated with BioProject [PRJNA692306](#) and deposited under SRA project [SRP301859](#) (SRA numbers [SRR13441373](#) to [SRR13441375](#)).

ACKNOWLEDGMENTS

J.E.S. is a CIFAR fellow in the program “Fungal Kingdom: Threats and Opportunities” and was supported by the U.S. Department of Agriculture, National Institute of Food and Agriculture Hatch project CA-R-PPA-211-5062-H and National Science Foundation (NSF) award DEB-1441715. H.S.J. was supported by NSF award IOS-1753749. Sequencing was supported in part by Cornell University startup funds (M.A.G.) and a grant from the New York State Department of Agriculture & Markets (C00237GG) to C.D.S.

Data analyses were performed on the high-performance computing cluster at the University of California—Riverside in the Institute of Integrative Genome Biology, supported by NSF DBI-1429826 and NIH S10-OD016290.

REFERENCES

1. Stam R, Jupe J, Howden AJM, Morris JA, Boevink PC, Hedley PE, Huitema E. 2013. Identification and characterisation CRN Effectors in *Phytophthora capsici* shows modularity and functional diversity. *PLoS One* 8:e59517. <https://doi.org/10.1371/journal.pone.0059517>.
2. Stam R, Howden AJM, Delgado-Cerezo M, Amaro TM, Motion GB, Pham J, Huitema E. 2013. Characterization of cell death inducing *Phytophthora capsici* CRN effectors suggests diverse activities in the host nucleus. *Front Plant Sci* 4:387. <https://doi.org/10.3389/fpls.2013.00387>.
3. Vega-Arreguin JC, Jalloh A, Bos JI, Moffett P. 2014. Recognition of an Avr3a homologue plays a major role in mediating nonhost resistance to *Phytophthora capsici* in *Nicotiana* species. *Mol Plant Microbe Interact* 27:770–780. <https://doi.org/10.1094/MPMI-01-14-0014-R>.
4. Mafurah JJ, Ma H, Zhang M, Xu J, He F, Ye T, Shen D, Chen Y, Rajput NA, Dou D. 2015. A virulence essential CRN effector of *Phytophthora capsici* suppresses host defense and induces cell death in plant nucleus. *PLoS One* 10:e0127965. <https://doi.org/10.1371/journal.pone.0127965>.
5. Li Q, Ai G, Shen D, Zou F, Wang J, Bai T, Chen Y, Li S, Zhang M, Jing M, Dou D. 2019. A *Phytophthora capsici* effector targets ACD11 binding partners that regulate ROS-mediated defense response in *Arabidopsis*. *Mol Plant* 12:565–581. <https://doi.org/10.1016/j.molp.2019.01.018>.
6. Chen X-R, Zhang Y, Li H-Y, Zhang Z-H, Sheng G-L, Li Y-P, Xing Y-P, Huang S-X, Tao H, Kuan T, Zhai Y, Ma W. 2019. The RXLR effector PcAvh1 is required for full virulence of *Phytophthora capsici*. *Mol Plant Microbe Interact* 32:986–1000. <https://doi.org/10.1094/MPMI-09-18-0251-R>.
7. Li Q, Wang J, Bai T, Zhang M, Jia Y, Shen D, Zhang M, Dou D. 2020. A *Phytophthora capsici* effector suppresses plant immunity via interaction with EDS1. *Mol Plant Pathol* 21:502–511. <https://doi.org/10.1111/mpp.12912>.
8. Lamour KH, Stam R, Jupe J, Huitema E. 2012. The oomycete broad-host-range pathogen *Phytophthora capsici*. *Mol Plant Pathol* 13:329–337. <https://doi.org/10.1111/j.1364-3703.2011.00754.x>.
9. Quesada-Ocampo LM, Granke LL, Mercier MR, Olsen J, Hausbeck MK. 2011. Investigating the genetic structure of *Phytophthora capsici* populations. *Phytopathology* 101:1061–1073. <https://doi.org/10.1094/PHYTO-11-10-0325>.
10. Parada-Rojas CH, Quesada-Ocampo LM. 2018. Analysis of microsatellites from transcriptome sequences of *Phytophthora capsici* and applications for population studies. *Sci Rep* 8:5194. <https://doi.org/10.1038/s41598-018-23438-8>.
11. Lamour KH, Mudge J, Gobena D, Hurtado-Gonzales OP, Schmutz J, Kuo A, Miller NA, Rice BJ, Raffaele S, Cano LM, Bharti AK, Donahoo RS, Finley S, Huitema E, Hulvey J, Platt D, Salamov A, Savidor A, Sharma R, Stam R, Storey D, Thines M, Win J, Haas BJ, Dinwiddie DL, Jenkins J, Knight JR, Affourtit JP, Han CS, Chertkov O, Lindquist EA, Detter C, Grigoriev IV, Kamoun S, Kingsmore SF. 2012. 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. <https://doi.org/10.1094/MPMI-02-12-0028-R>.
12. Chen X-R, Xing Y-P, Li Y-P, Tong Y-H, Xu J-Y. 2013. RNA-Seq reveals infection-related gene expression changes in *Phytophthora capsici*. *PLoS One* 8:e74588. <https://doi.org/10.1371/journal.pone.0074588>.
13. Lamour K, Finley L. 2006. A strategy for recovering high quality genomic DNA from a large number of *Phytophthora* isolates. *Mycologia* 98:514–517. <https://doi.org/10.3852/mycologia.98.3.514>.
14. Kim KE, Peluso P, Babayan P, Yeadon PJ, Yu C, Fisher WW, Chin C-S, Rapicavoli NA, Rank DR, Li J, Catchside DEA, Celniker SE, Phillippy AM, Bergman CM, Landolin JM. 2014. Long-read, whole-genome shotgun sequence data for five model organisms. *Sci Data* 1:140045. <https://doi.org/10.1038/sdata.2014.45>.
15. Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30:3506–3514. <https://doi.org/10.1093/bioinformatics/btu538>.
16. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>.
17. Zimin AV, Puiu D, Luo M-C, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J, Salzberg SL. 2017. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res* 27:787–792. <https://doi.org/10.1101/gr.213405.116>.
18. Stajich JE, Palmer J. 2019. AAFTE: v0.2.3: automatic assembly for the fungi. <https://doi.org/10.5281/zenodo.1620526>.
19. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
20. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548. <https://doi.org/10.1093/molbev/msx319>.
21. Palmer JM, Stajich JE. 2020. Funannotate v1.8.1: eukaryotic genome annotation. <https://doi.org/10.5281/zenodo.1134477>.
22. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 37:907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
23. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>.
24. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512. <https://doi.org/10.1038/nprot.2013.084>.
25. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9:R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
26. Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>.

27. Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syn-
tenically mapped cDNA alignments to improve *de novo* gene finding. *Bio-*
informatics 24:637–644. <https://doi.org/10.1093/bioinformatics/btn013>.
28. Ter-Hovhannisyantsyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008.
Gene prediction in novel fungal genomes using an *ab initio* algorithm
with unsupervised training. *Genome Res* 18:1979–1990. <https://doi.org/10.1101/gr.081612.108>.
29. Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM:
two open source *ab initio* eukaryotic gene-finders. *Bioinformatics*
20:2878–2879. <https://doi.org/10.1093/bioinformatics/bth315>.
30. Testa AC, Hane JK, Ellwood SR, Oliver RP. 2015. CodingQuarry: highly accurate
hidden Markov model gene prediction in fungal genomes using RNA-seq tran-
scripts. *BMC Genomics* 16:170. <https://doi.org/10.1186/s12864-015-1344-4>.
31. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment
using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>.
32. Slater GSC, Birney E. 2005. Automated generation of heuristics for biologi-
cal sequence comparison. *BMC Bioinformatics* 6:31. <https://doi.org/10.1186/1471-2105-6-31>.
33. UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge.
Nucleic Acids Res 47:D506–D515. <https://doi.org/10.1093/nar/gky1049>.
34. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang H-Y,
Dosztányi Z, El-Gebali S, Fraser M, Gough J, Haft D, Holliday GL, Huang H,
Huang X, Letunic I, Lopez R, Lu S, Marchler-Bauer A, Mi H, Mistry J, Natale DA,
Necci M, Nuka G, Orengo CA, Park Y, Pesseat S, Piovesan D, Potter SC,
Rawlings ND, Redaschi N, Richardson L, Rivoire C, Sangrador-Vegas A, Sigrist
C, Sillitoe I, Smithers B, Squizzato S, Sutton G, Thanki N, Thomas PD, Tosatto
SCE, Wu CH, Xenarios I, Yeh L-S, Young S-Y, Mitchell AL. 2017. InterPro in
2017: beyond protein family and domain annotations. *Nucleic Acids Res* 45:
D190–D199. <https://doi.org/10.1093/nar/gkw1107>.
35. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von
Mering C, Bork P. 2017. Fast genome-wide functional annotation through
orthology assignment by eggNOG-Mapper. *Mol Biol Evol* 34:2115–2122.
<https://doi.org/10.1093/molbev/msx148>.
36. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y.
2018. dbCAN2: a meta server for automated carbohydrate-active enzyme
annotation. *Nucleic Acids Res* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>.
37. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. 2018.
The MEROPS database of proteolytic enzymes, their substrates and inhibi-
tors in 2017 and a comparison with peptidases in the PANTHER database.
Nucleic Acids Res 46:D624–D632. <https://doi.org/10.1093/nar/gkx1134>.