



Whole-Genome Sequence Data for the Holotype Strain of *Diaporthe ilicicola*, a Fungus Associated with Latent Fruit Rot in Deciduous Holly

Isabel B. Emanuel,^a Zachary M. Konkel,^b Kelsey L. Scott,^a Guillermo E. Valero David,^a  Jason C. Slot,^a  Francesca Peduto Hand^a

^aDepartment of Plant Pathology, The Ohio State University, Columbus, Ohio, USA

^bCenter for Applied Plant Sciences, The Ohio State University, Columbus, Ohio, USA

ABSTRACT *Diaporthe ilicicola* is a newly described fungal species that is associated with latent fruit rot in deciduous holly. This announcement provides a whole-genome assembly and annotation for this plant pathogen, which will inform research on its parasitism and identification of gene clusters involved in the production of bioactive metabolites.

D*iaporthe* species (*Phomopsis*; anamorph) are ascomycete fungi found as saprophytes, endophytes, and plant pathogens (1) and are recognized for their diverse production of metabolites (2). This report announces the whole-genome assembly and annotation of *Diaporthe ilicicola*, a newly described species associated with fruit rot in winterberry (*Ilex* spp.). Data will inform research to understand how plant-pathogenic *Diaporthe* species differ from those that do not cause disease and will aid in the identification of fungal gene clusters involved in bioactive metabolite production.

Diaporthe ilicicola holotype strain FPH2015-502 (CBS 144318) was isolated as described by Lin et al. (3). A single spore culture was maintained in long-term storage at -20°C . *Diaporthe ilicicola* FPH2015-502 was grown in potato dextrose broth for 2 days at 28°C and 150 rpm. DNA was extracted with the DNeasy Plant minikit (Qiagen) according to the manufacturer's recommendations and sequenced with both short- and long-read sequencing. Illumina library preparation and sequencing were conducted at Novogene, Inc., using a NovaSeq 6000 system with paired-end 150-bp reads. The long-read sequencing library was prepared using the Oxford Nanopore Technologies SQK-LSK108 kit under standard preparation conditions with g-TUBE fragmentation. Long reads were sequenced using a Nanopore MinION Mk1B flow cell with R9.4.1 chemistry.

For genome assembly and annotation, default parameters were used for all software unless specified otherwise. Illumina read quality was checked using FastQC v0.11.8, and reads were trimmed using Trimmomatic v0.36 (options: HEADCROP:10 CROP:145 SLIDINGWINDOW:50:25 MINLEN:100) (4) for a total of 24,778,254 paired-end reads. Nanopore data were base called with Guppy v2.1.3 using the baseline model (Oxford Nanopore Technologies); reads were quality filtered and adapters were removed with Porechop v0.2.4 (5) for total of 136,483 reads (total bases, 0.62 Gbp; read N_{50} , 7,501 bp) (Table 1). Hybrid genome assembly was performed using SPAdes v3.12.0 (kmer lengths, 21, 33, 55, 77, and 99) (6), FastQC v0.11.8 (7) was used to check assembly quality, and QUAST v4.6.3 (8) was used to compare the assembly with one reference organism, *Diaporthe helianthi* strain 7/96 (GenBank accession number [MAVT000000000](#)) (9). Repetitive elements were masked using RepeatModeler v1.0.11 (10) referencing the Dfam 3.0 repeat library (11). The assembly was annotated using the Funannotate pipeline v1.8.1 (12) with the following associated software: GlimmerHMM v3.0.4 (13), GeneMark-ES v4.35 (14), SNAP v2006-07-28 (15), BUSCO v3.0.2 (16), and AUGUSTUS v3.3.3 (17). Protein evidence was supplied from seven *Sordariomycetes* species, including *Diaporthe ampelina* UCDDA912 (GenBank accession number [LCUC000000000](#)) (18), *D. helianthi* 7/96 strain (GenBank accession number [MAVT000000000](#)) (9), *Cytospora*

Editor Vincent Bruno, University of Maryland School of Medicine

Copyright © 2022 Emanuel et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

Address correspondence to Francesca Peduto Hand, hand.81@osu.edu.

The authors declare no conflict of interest.

Received 1 July 2022

Accepted 5 August 2022

Published 22 August 2022

TABLE 1 Assembly and annotation statistics for *D. ilicicola* FPH2015-502, comparison species *D. helianthi*, and *D. ampelina*

Parameter	Data for:		
	<i>Diaporthe ilicicola</i> (GenBank accession no. JALPVH000000000)	<i>Diaporthe ampelina</i> (GenBank accession no. LWAD000000000) ^a	<i>Diaporthe helianthi</i> (GenBank accession no. MAVT000000000) ^b
Assembly			
Genome size (bp)	65,225,202	47,325,858	63,672,038
No. of contigs	1,399	2,383	7,376
N_{50} (bp)	156,399	132,346	20,184
Contig L_{50}	123	103	860
GC content (%)	44.12	53.94	43.99
Complete BUSCOs (%) ^c	96.3	96.3	96.4
Annotation			
No. of predicted genes	12,061	10,704	13,139
Mean gene length (bp)	1,672	1,535	1,615
Proportion of assembly covered by annotation (%)	30.92	34.71	33.32

^a From Savitha et al. (21). The genome of *D. ampelina* is included for a robust comparison of assembly and annotation quality across *Diaporthe* spp.

^b From Baroncelli et al. (9).

^c The completeness (percentage of complete benchmarking universal single-copy orthologs [BUSCOs]) of the genomes was assessed using BUSCO v3.0.2 and the sordariomyceta database (release 9).

leucostoma (BioSample accession number [SAMN04099706](#)), *Valsa mali* (BioSample accession number [SAMN03203459](#)), *Valsa malicola* (BioSample accession number [SAMN04099704](#)), *Valsa mali* var. *pyri* (BioSample accession number [SAMN03203462](#)), and *Valsa sordida* (BioSample accession number [SAMN04099705](#)). The annotation was finalized using a modified version of OrthoFiller v1.2xonq (19) (<https://gitlab.com/xonq/orthofiller>) referencing 15 *Diaporthales* species, namely, *Coniella lustricola* (<https://mycocosm.jgi.doe.gov/Pilidi1/Pilidi1.home.html>), *Cryphonectria parasitica* (BioSample accession number [SAMN02744051](#)), *Cryptodiaporthe* sp. (<https://mycocosm.jgi.doe.gov/Crypto1/Crypto1.home.html>), *Cryptodiaporthe populea* (<https://mycocosm.jgi.doe.gov/Crypto1/Crypto1.home.html>), *Cytospora chrysosperma* (<https://mycocosm.jgi.doe.gov/Cytc1/Cytc1.home.html>), *Cytospora leucostoma* (BioSample accession number [SAMN04099706](#)), *Diaporthe ampelina* (<https://mycocosm.jgi.doe.gov/Diaam1/Diaam1.home.html>) (18), *Diaporthe citri* (BioSample accession number [SAMN15772025](#)), *Diaporthe helianthi* (<https://mycocosm.jgi.doe.gov/Diahe1/Diahe1.home.html>) (9), *Diaporthaceae* sp. (https://mycocosm.jgi.doe.gov/DiaPMI573_1/DiaPMI573_1.home.html), *Lollipopaia minuta* (<https://mycocosm.jgi.doe.gov/Lolmi1/Lolmi1.home.html>), *Melanconium* sp. (<https://mycocosm.jgi.doe.gov/Melsp1/Melsp1.home.html>), *Valsa mali* (<https://mycocosm.jgi.doe.gov/Valma1/Valma1.home.html>) (20), *Valsa mali* (BioSample accession number [SAMN03203462](#)), and *Valsa sordida* (BioSample accession number [SAMN04099705](#)).

Data availability. The genome was deposited in DDBJ/ENA/GenBank under the accession number [JALPVH000000000](#) (BioProject accession number [PRJNA759853](#), BioSample accession number [SAMN21205738](#), and SRA accession numbers [SRR18821933](#) and [SRR18821934](#)).

ACKNOWLEDGMENTS

The research described in this paper fulfills a portion of the requirements for the Ph.D. degree in Plant Pathology at the Ohio State University for I.B.E.

Funding was provided by USDA-NIFA Hatch Project 1020446 to F.P.H. and J.C.S. and National Science Foundation grant DEB-1638999 to J.C.S. None of the funding sources had a role in study design, data collection, or data interpretation.

REFERENCES

- Gomes RR, Glienke C, Videira SIR, Lombard L, Groenewald JZ, Crous PW. 2013. *Diaporthe*: a genus of endophytic, saprobic and plant pathogenic fungi. *Persoonia* 31:1–41. <https://doi.org/10.3767/003158513X666844>.
- Chepkirui C, Stadler M. 2017. The genus *Diaporthe*: a rich source of diverse and bioactive metabolites. *Mycol Prog* 16:477–494. <https://doi.org/10.1007/s11557-017-1288-y>.
- Lin S, Taylor NJ, Peduto Hand F. 2018. Identification and characterization of fungal pathogens causing fruit rot of deciduous holly. *Plant Dis* 102:2430–2445. <https://doi.org/10.1094/PDIS-02-18-0372-RE>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.

5. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20:129. <https://doi.org/10.1186/s13059-019-1727-y>.
6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19: 455–477. <https://doi.org/10.1089/cmb.2012.0021>.
7. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
8. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
9. Baroncelli R, Scala F, Vergara M, Thon MR, Ruocco M. 2016. Draft whole-genome sequence of the *Diaporthe helianthi* 7/96 strain, causal agent of sunflower stem canker. *Genom Data* 10:151–152. <https://doi.org/10.1016/j.gdata.2016.11.005>.
10. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* 117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
11. Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, Smit AFA, Finn RD. 2013. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res* 41:D70–D82. <https://doi.org/10.1093/nar/gks1265>.
12. Palmer J, Stajich J. 2020. Funannotate v1.8.1: eukaryotic genome annotation. Zenodo <https://doi.org/10.5281/zenodo.1134477>.
13. 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>.
14. Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* 26:1107–1115. <https://doi.org/10.1093/nar/26.4.1107>.
15. Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>.
16. 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>.
17. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 32:W309–W312. <https://doi.org/10.1093/nar/gkh379>.
18. Morales-Cruz A, Amrine KC, Blanco-Ulate B, Lawrence DP, Travadon R, Rolshausen PE, Baumgartner K, Cantu D. 2015. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genomics* 16:469. <https://doi.org/10.1186/s12864-015-1624-z>.
19. Dunne MP, Kelly S. 2017. OrthoFiller: utilising data from multiple species to improve the completeness of genome annotations. *BMC Genomics* 18: 390. <https://doi.org/10.1186/s12864-017-3771-x>.
20. Yin Z, Liu H, Li Z, Ke X, Dou D, Gao X, Song N, Dai Q, Wu Y, Xu JR, Kang Z, Huang L. 2015. Genome sequence of *Valsa* canker pathogens uncovers a potential adaptation of colonization of woody bark. *New Phytol* 208:1202–1216. <https://doi.org/10.1111/nph.13544>.
21. Savitha J, Bhargavi SD, Praveen VK. 2016. Complete genome sequence of the endophytic fungus *Diaporthe (Phomopsis) ampelina*. *Genome Announc* 4:e00477–16. <https://doi.org/10.1128/genomeA.00477-16>.