High-Quality Genome Resource of the Pathogen of *Diaporthe (Phomopsis) phragmitis* Causing Kiwifruit Soft Rot

Xiaojie Wang,^{1,3} Hongjie Dong,¹ Jianbin Lan,² Yuanyuan Liu,¹ Kuan Liang,¹ Qi Lu,¹ Zemin Fang,³ and Pu Liu^{1,†}

¹ School of Horticulture, Anhui Agricultural University, Hefei 230036, People's Republic of China

² College of Landscape Architecture and Life Science, Chongqing University of Arts and Sciences, Chongqing 402160, China

³ School of Life Sciences, Anhui University, Hefei 230601, China

Abstract

Diaporthe spp. are critical plant pathogens that cause wood cankers, wilt, dieback, and fruit rot in a wide variety of economic plant hosts and are regarded as one of the most acute threats faced by the kiwifruit industry worldwide. *Diaporthe phragmitis* NJD1 is a highly pathogenic isolate of soft rot of kiwifruit. Here, we present a high-quality genome-wide sequence of *D. phragmitis* NJD1 that was assembled into 28 contigs containing a total size of 58.33 Mb and N₅₀ length of 3.55 Mb. These results lay a solid foundation for understanding host-pathogen interaction and improving disease management strategies.

Funding

This project was supported by the National Key Research and Development Program of China (2018YFD1000300); Anhui Key Research and Development Program (1804a07020105 and 201904f06020035) and Natural Science Foundation of Anhui Province of China (2008085MC78); National Natural Foundation of China grant 32072548.

Keywords

Diaporthe phragmitis, high virulence, kiwifruit, soft rot

Genome Announcement

Kiwifruit is widely planted and consumed for its unique flavor and nutritional value. China is the largest kiwifruit producer in the world. The cultivation area is more than 168,000 ha in 2018, with a productivity of 2,035,158 tons (Food and Agriculture Organization of the United Nations Faostat). Kiwifruit softening and decay caused by soft rot are the key deterioration factors in the postharvest storage process (Kinugawa and Sato 2003). In previous studies, the species of the genus *Diaporthe*, including *D. actinidiae* (Lee et al. 2001; Pintos-Varela et al. 2000; Sommer and Beraha 1975), *D. ambigua* (Auger et al. 2013), *D. australafricana* (Luongo et al. 2011), *D. lithocarpus* (Li et al. 2016a and b), *D. novem* (Díaz et al. 2014), *D. perniciososa*, *D. rudis* (Díaz et al. 2017), and *D. viticola* (Luongo et al. 2011), have been described causing soft rots and decays on kiwifruit. The typical symptoms include soft rot with brown skin that starts at the stem end or middle of the fruit; in more serious instances, it will affect the whole kiwifruit. From the inside, affected fruits are obviously watery and disorganized and are then severely decayed and sour smelling.

In 2019, *D. phragmitis* NJD1 was isolated from the soft rot kiwifruit fruit of *Actinidia chinensis* cultivar Donghong in Jinzhai County, Anhui Province, P.R China. The fruit samples

[†]Corresponding author: P. Liu; puliu@ahau.edu.cn

Xiaojie Wang, Hongjie Dong, and Jianbin Lan are co-first authors.

*The *e-Xtra* logo stands for “electronic extra” and indicates there is supplementary material published online.

The author(s) declare no conflict of interest.

Accepted for publication 21 October 2020.

Copyright © 2021 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

Table 1. General features of the genome

Feature	<i>Diaporthe phragmitis</i>
Total genome size (bp)	58,328,132
Contigs	28
N ₅₀ length (bp)	3,550,333
GC (%)	50.82
Gene number	12,393
Gene average length (bp)	1,317
Gene length/genome (%)	27.98
BUSCO	
Complete	97.9%
Missing	1.1%
Complete and duplicated BUSCOs	1.0%

were surface-sterilized by an aqueous solution of NaClO (1% active chlorine) in 50% alcohol. Pathogenicity assessment showed that NJD1 is a high-virulence strain in kiwifruit. Based on internal transcribed spacer, β -tubulin, and histone H3 sequences, the species was identified as *D. phragmitis*. The genomic DNA of strain NJD1 was extracted from a fresh mycelial suspension after 3 days of culture in potato dextrose broth medium at 25°C by using the cetyl trimethyl ammonium bromide (CTAB) method (Weiland 1997). The whole-genome sequencing of NJD1 was performed with PacBio long-read and Illumina short-read sequencing technologies at the Beijing Novogene Bioinformatics Technology Co., Ltd. (Beijing). The 350-bp paired-end library was prepared separately, following Illumina protocol, and was sequenced using the Illumina NovaSeq PE150 (coverage 50×). The libraries for single-molecule real-time (SMRT) sequencing were separately constructed, using the PacBio Sequel platform, and the insert size was 20 kb.

SPAdes runs at many different k-mer levels (35, 47, 59, 71, 83, 95, 107, 119) were conducted to obtain the best assembly. Finally, the whole-genome size was approximately 56.29 Mb (k-mer depth 15.81) with a heterozygous rate of 0.08% (Vurture et al. 2017). The integrity of the total genome assembly was assessed with BUSCO version 3.0.2 (Simão et al. 2015) against fungi_odb9. After the low-quality reads were filtered, the remaining high-quality reads were finally assembled into contigs by SMRT Link v5.0.1 (Ardui et al. 2018; Reiner et al. 2018). A total of 555,967 subreads with an average read length of 11,833 bp and a sub-read N₅₀ of 13,409 bp were generated for *D. phragmitis* NJD1. The reads were assembled into 28 contigs with a N₅₀ of 3.55 Mb and a maximum contig length of 7.71 Mb. The size of the draft genome was 58.3 Mb, with a G+C content of 50.82% (Table 1). The automatic error correction function of the SMRT portal were used to improve the accuracy of the seed sequence. Genes in the assembly were predicted using Augustus (Stanke et al. 2006), SNAP (Korf 2004), and GeneMarkhm (Lukashin and Borodovsky 1998). PASA (Program to Assemble Spliced Alignments) alignment assemblies based on overlapping transcript alignments from step and use of EVIDENCEModeler to compute weighted consensus gene structure annotations (Haas et al. 2008). A total of 12,393 genes were identified that encode proteins, with an average length of 1,317 bp, accounting for 27.98% of the whole genome (Table 1). Furthermore, ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs), and microRNAs (miRNAs) were predicted and analyzed by RNAmmer (Lagesen et al. 2007), tRNAscan-SE (Lowe and Chan 2016), and Rfam database (Nawrocki et al. 2015). The set of the predicted noncoding genes included 37 rRNAs, 174 tRNAs, 16 snRNAs, two sRNAs, and without miRNA. Interspersed repetitive sequences were predicted and were annotated using RepeatMasker (Smit et al. 2016). The proportion of tandem repeat rate was 2.12%, comprising 880 long terminal repeats, 512 long interspersed nuclear elements, and 641 microsatellite DNA.

A genome-wide BLAST search (e-value < 1e-5, the minimum alignment length percentage >40%) was performed on the protein database to annotate the *D. phragmitis* NJD1 genome. Of the predicted coding genes, 11,624, 7,853, 10,207, and 2,206 genes had homologs with known functions in the Non-Redundant Protein, Gene Ontology, Kyoto Encyclopedia of Genes and Genomes, and Clusters of Orthologous Groups databases. For *D. phragmitis* NJD 1, as a plant fungal pathogen, possible genes related to pathogenicity were analyzed. A total of 1,451 proteins were predicted to be related to pathogen-host interactions,

and 558 proteins were considered to act as virulence factors, based on the best-matched proteins from the Pathogen Host Interactions database (Urban et al. 2015) and the database of fungal virulence factors (Lu et al. 2012). A total of 806 proteins were categorized as carbohydrate-active enzymes in the Carbohydrate-Active Enzymes database (Lombard et al. 2014), of which 404, 127, and 37 were identified as glycoside hydrolases, glycosyl transferases, and polysaccharide lyases, respectively (Saier et al. 2016).

Secretory proteins of fungal pathogens, such as effectors, play important roles in their pathogenicity and virulence. The secretory proteins were analyzed with the Signal P database (version 4.1) for the signal peptide and with TMHMM (version 2.0c) for the transmembrane domains (Petersen et al. 2011). The proteins without transmembrane domains or only one transmembrane domain were further analyzed with WoLF PSORT v0.2 software for their subcellular localization (Horton et al. 2007). Proteins were classified as secretomes when their predicted extracellular score was >15. A total of 1,172 credible secretory proteins were predicted. Secondary metabolic products, such as mycotoxins, have been identified as critical virulence factors of fungal pathogens in previous reports (Goswami and Kistler 2005; Proctor et al. 1995). Therefore, putative genes and clusters related to secondary metabolism were predicted by the web-based tool antiSMASH 2.0 (Blin et al. 2013). Finally, 806 genes were identified in 80 clusters associated with secondary metabolites, including 37 t1 polyketide synthase (PKS) clusters (403 genes), three nonribosomal peptide synthetase (NRPS) clusters (134 genes), three terpene clusters (88 genes), one PKS-NRPS cluster (35 genes), and 125 genes in other clusters.

The genomes of *D. phragmitis* NJD1 were uploaded to GenBank under the accession numbers JACDXY000000000 (BioSample SAMN15232972) and the annotation information is provided in Appendix S1 in the Supplementary Material.

Author-Recommended Internet Resources

Carbohydrate-Active Enzymes database: <http://www.cazy.org>

Food and Agriculture Organization of the United Nations FAOSTAT: <http://www.fao.org/faostat/en>

RepeatMasker: <http://www.repeatmasker.org>

Literature Cited

- Ardui, S., Ameer, A., Vermeesch, J. R., and Hestand, M. S. 2018. Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Res.* 46:2159-2168.
- Auger, J., Pérez, I., and Esterio, M. 2013. *Diaporthe ambigua* associated with post-harvest fruit rot of kiwifruit in Chile. *Plant Dis.* 97:843.
- Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. 2013. antiSMASH 2.0—A versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* 41:W204-W212.
- Díaz, G. A., Latorre, B. A., Jara, S., Ferrada, E., Naranjo, P., Rodríguez, J., and Zoffoli, J. P. 2014. First report of *Diaporthe novem* causing postharvest rot of kiwifruit during controlled atmosphere storage in Chile. *Plant Dis.* 98: 1274.
- Díaz, G. A., Latorre, B. A., Lolas, M., Ferrada, E., Naranjo, P., and Zoffoli, J. P. 2017. Identification and characterization of *Diaporthe ambigua*, *D. australifrica*, *D. novem*, and *D. rudis* causing a postharvest fruit rot in kiwifruit. *Plant Dis.* 101: 1402-1410.
- Goswami, R. S., and Kistler, H. C. 2005. Pathogenicity and in planta mycotoxin accumulation among members of the *Fusarium graminearum* species complex on wheat and rice. *Phytopathology* 95:1397-1404.
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., and Wortman, J. R. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. *Genome Biol.* 9:R7.
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., and Nakai, K. 2007. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* 35 (Web Server):W585-W587.
- Kinugawa, M., and Sato, T. 2003. Kiwifruit dieback caused by *Diaporthe* sp. and *Botryosphaeria dothidea*, kiwifruit soft rot fungi. *Ann. Phytopathol. Soc. Jpn.* 69: 373-383.
- Korf, I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. 2007. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100-3108.
- Lee, J. G., Lee, D.-H., Park, S.-Y., Hur, J.-S., and Koh, Y.-J. 2001. First report of *Diaporthe actinidiae*, the causal organism of stem-end rot of kiwifruit in Korea. *Plant Pathol. J.* 17:110-113.
- Li, L., Chen, M. Y., Zhang, P., Han, F., Liu, X. L., and Zhong, C. H. 2016a. Identification of the pathogen causing fruit soft rot on kiwifruit. *J. Plant Prot.* 43: 527-528.
- Li, L., Pan, H., Chen, M. Y., and Zhong, C. H. 2016b. First report of *Diaporthe lithocarpus* causing postharvest rot of kiwifruit in Sichuan province, China. *Plant Dis.* 100:2327.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42 (D1):D490-D495.
- Lowe, T. M., and Chan, P. P. 2016. tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44 (W1): W54-W57.
- Lu, T., Yao, B., and Zhang, C. 2012. DFVF: Database of fungal virulence factors. Database (Oxford) 2012:bas032.
- Lukashin, A. V., and Borodovsky, M. 1998. GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res.* 26:1107-1115.
- Luongo, L., Santori, A., Riccioni, L., and Belisario, A. 2011. *Phomopsis* sp. associated with post-harvest fruit rot of kiwifruit in Italy. *J. Plant Pathol.* 93:205-209.
- Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., Floden, E. W., Gardner, P. P., Jones, T. A., Tate, J., and Finn, R. D. 2015. Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.* 43 (D1):D130-D137.
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785-786.

- Pintos-Varela, C., García-Jiménez, J., Mansilla, J. P., Ciurana, N., Sales, R., and Armengol, J. 2000. Presencia de *Diaporthe actinidiae* afectando al kiwi (*Actinidia deliciosa*) en el noroeste de la península Ibérica. Bol. San. Veg. 26:389-399 (In Spanish).
- Proctor, R. H., Hohn, T. M., and McCormick, S. P. 1995. Reduced virulence of *Gibberella zeae* caused by disruption of a trichothecene toxin biosynthetic gene. Mol. Plant-Microbe Interact. 8:593-601.
- Reiner, J., Pisani, L., Qiao, W., Singh, R., Yang, Y., Shi, L., Khan, W. A., Sebra, R., Cohen, N., Babu, A., Edelmann, L., Jabs, E. W., and Scott, S. A. 2018. Cytogenomic identification and long-read single molecule real-time (SMRT) sequencing of a *Bardet-Biedl Syndrome 9* (*BBS9*) deletion. NPJ Genom. Med. 3:3.
- Saier, M. H., Jr., Reddy, V. S., Tsu, B. V., Ahmed, M. S., Li, C., and Moreno-Hagelsieb, G. 2016. The transporter classification database (TCDB): Recent advances. Nucleic Acids Res. 44 (D1):D372-D379.
- 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.
- Smit, A. F. A., Hubley, R., and Green, P. 2016. Masker Open-4.0. 2013-2015. Institute for Systems Biology, Seattle, WA, U.S.A.
- Sommer, N. F., and Beraha, L. 1975. *Diaporthe actinidiae*, a new species causing stem-end rot of Chinese gooseberry fruits. Mycologia 67:650-653.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. 2006. AUGUSTUS: Ab initio prediction of alternative transcripts. Nucleic Acids Res. 34: W435-W439.
- Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H., and Hammond-Kosack, K. E. 2015. The Pathogen-Host Interactions database (PHI-base): Additions and future developments. Nucleic Acids Res. 43 (D1):D645-D655.
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., and Schatz, M. C. J. B. 2017. GenomeScope: Fast reference-free genome profiling from short reads. Bioinformatics 33:2202-2204.
- Weiland, J. J. 1997. Rapid procedure for the extraction of DNA from fungal spores and mycelia. Fungal Genet. Newsl. 44:60-63.