

Whole-genome sequence and annotation of *Diaporthe australafricana* strain CBS 118571

Xin Li,¹ Xinyi Wang,² Limei Li,³ Rong Lei⁴

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the whole-genome sequence of *Diaporthe australafricana* Crous & J.M. van Niekerk using Oxford Nanopore long-read sequencing and Illumina short-read sequencing. The hybrid genome consists of 11 contigs with a total length of 53.509 Mb, and a GC content of 52.40%.

KEYWORDS whole-genome sequence, annotation, *Diaporthe australafricana*

Diaporthe australafricana Crous & J.M. van Niekerk strain CBS 135771 was originally isolated from grapevines in the Western Cape province, Stellenbosch in South Africa (1), and deposited at CBS-KNAW, Utrecht, Netherlands. *D. australafricana* is highly virulent in shoots, stems, and fruits of blueberry (2), causing stem canker and dieback, lesions on stems, and necrosis of shoots (3). Here, we performed *de novo* assembly and annotation of *D. australafricana* genome using Oxford Nanopore long-read sequencing and Illumina short-read sequencing.

The white and fluffy aerial mycelium colonies were cultivated on a PDA culture medium, and grown at 20°C under a 12-h dark/12-h light cycle for approximately 6 days. The mycelia were collected and ground with liquid nitrogen, then the genomic DNA was extracted using a DNeasy Plant Pro Kit (Qiagen, GmbH, Hilden, Germany), and the total RNA was extracted using an RNeasy Pure Plant Plus Kit (Tiangen, Beijing, China). The RNA integrity number was analyzed with an Agilent RNA 6000 Nano Kit on an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). To construct the ONT DNA library, genomic DNA was fragmented with a g-Tube (Cat No. 520079, Covaris, USA) and the long DNA fragments were retained with SPRIselect Reagent (Cat No. B23318, Beckman, USA). Next, NEBNext FFPE Repair Mix (Cat# M6630) and NEBNext Ultra II End Repair/dA-tailing Kit (Cat# E7546) were used to end-repair and A-tailing for the fragmented DNA. The adapter in the SQK-LSK109 (ONT, UK) was mixed with NEBNext Quick Ligation Reaction Buffer (Cat# B6058) and NEB T4 DNA Ligase 2 M U/mL (Cat# M0202) for ligation. Approximately 600 ng library was loaded to a MinION Flow Cell (R9.4.1) in an Oxford Nanopore promethION P48 platform (Oxford Nanopore Technologies, UK) for sequencing. For Illumina sequencing, 1 µg genomic DNA was fragmented using NEBNext dsDNA Fragmentase (Cat# M0348), and library was constructed using the ExCell Bio Library Construction Kit (NuGEN, San Carlos, CA, USA). Illumina RNA-Seq library was constructed with the TransNGS Stranded RNA-Seq Library Prep Kit (TransGen, Beijing, China). The quality control of libraries was performed on an Agilent 2100 Bioanalyzer, and the libraries were sequenced on an Illumina NovaSeq 6000 platform (San Diego, CA, USA). For Nanopore sequencing, the basecalling from the raw FAST5 files was performed using guppy (v6.5.7) (4) with the basecalling model of dna_r9.4.1_450bps_hac, followed by adapter trimming using PoreChop (v0.2.4) (5). For Illumina sequencing, the raw data (12,453,433,200 bp with PE150 and 893,925,500 bp with PE250) were trimmed of adapters and low-quality reads (mean quality value <20) by using skewer (6) (v0.2.2). Default parameters were used for the software except where otherwise noted.

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Rong Lei, leir@caiq.org.cn.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 23 May 2024

Accepted 7 August 2024

Published 30 August 2024

Copyright © 2024 Li 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/).

TABLE 1 Genome feature of *D. australafricana* CBS 135771

Features	Strain CBS135771
ONT long reads (sequencing depth)	7.748 Gb (~144.8×)
Average read length	18,252.08
Reads N_{50}	26,242
Maximum read length	135,101
Illumina short reads (sequencing depth)	13.35 Gb (~249.4×)
Estimated genome size	52.90 Mb
Estimated genome repeats	6.06%
Estimated genome heterozygosity	0.08%
RNA-Seq	6.31 Gb (~117.9×)
Assemble size	53.51 Mb
Contig number	11
Contig N_{50}	5,751,947
Contig L_{50}	4
Average contig length	4,864,461.45
Maximum contig length	11,543,331
GC content	52.40%
Repeat sequence	6.60%
BUSCO fungal integrity	99.0%
Illumina reads mapping rate	99.56%
ONT long reads mapping rate	99.07%

The hybrid assembly of Illumina data sets and Nanopore long reads into contigs was performed using MaSuRCA (7) (v3.3.7). The completely assembled *D. australafricana* CBS 135771 contained 11 contigs with a cumulative size of 53.509 Mb and an N_{50} length of 5.75 Mb. The concordance and the integrity of the assembly were determined using Burrows-Wheeler Alignment tool (BWA, version: 0.7.13-r1126) (8) and BUSCO (9, 10) (v4.1.4), respectively, and the genome mapping ratios were 99.56% and 99.07%, respectively. All metrics regarding assembly are listed in Table 1.

The assembled genome was annotated using pipeline named funannotate (11) (v1.8.3), in which a genome-guided assembly process of RNA-Seq reads by Trinity (12) was invoked. Finally, genome annotation predicted annotated 15,235 protein-coding genes.

AUTHOR AFFILIATIONS

¹Technology Center of Dalian Customs District, Dalian, China

²School of Life and Health, Dalian University, Dalian, China

³Jilin Provincial Academy of Forestry Science, Changchun, China

⁴Chinese Academy of Inspection and Quarantine, Beijing, China

AUTHOR ORCIDs

Xin Li  <http://orcid.org/0000-0001-8077-4206>

Xinyi Wang  <http://orcid.org/0000-0002-7389-6018>

Limei Li  <http://orcid.org/0000-0002-4170-523X>

Rong Lei  <http://orcid.org/0000-0002-4043-9558>

FUNDING

Funder	Grant(s)	Author(s)
MOST National Key Research and Development Program of China (NKPs)	2021YFD1400100, 2021YFD1400103	Xin Li Rong Lei

AUTHOR CONTRIBUTIONS

Xin Li, Funding acquisition, Resources | Xinyi Wang, Validation, Writing – review and editing | Limei Li, Validation | Rong Lei, Investigation, Methodology, Writing – original draft

DATA AVAILABILITY

The whole-genome data of *D. australaficana* CBS 135771 presented here have been deposited at GenBank under the accession number [JBCEDM000000000](#). The accession numbers of BioProject and BioSample are [PRJNA1091203](#), and [SAMN 40589631](#), respectively. The SRA accession number of ONT reads, Illumina PE150 reads, Illumina PE250 reads, and RNA-Seq PE150 reads are [SRR29120074](#), [SRR29120077](#), [SRR29120076](#) and [SRR29120075](#), respectively.

REFERENCES

1. van Niekerk JM, Groenewald JZ, Farr DF, Fourie PH, Halleen F, Crous PW. 2005. Reassessment of *Phomopsis* species on grapevines. *Austral Plant Pathol* 34:27–39. <https://doi.org/10.1071/AP04072>
2. Elfar K, Torres R, Díaz GA, Latorre BA. 2013. Characterization of *Diaporthe australaficana* and *Diaporthe* spp. associated with stem canker of blueberry in Chile. *Plant Dis* 97:1042–1050. <https://doi.org/10.1094/PDIS-11-12-1030-RE>
3. Latorre BA, Elfar K, Espinoza JG, Torres R, Díaz GA. 2012. First report of *Diaporthe australaficana* associated with stem canker on blueberry in Chile. *Plant Dis* 96:768. <https://doi.org/10.1094/PDIS-12-11-1025-PDN>
4. guppy (v6.5.7). https://cdn.oxfordnanoportal.com/software/analysis/ont-guppy_6.5.7_linux64.tar.gz.
5. PoreChop (v0.2.4). <https://github.com/rrwick/Porechop>.
6. Jiang H, Lei R, Ding SW, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15:182. <https://doi.org/10.1186/1471-2105-15-182>
7. 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>
8. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
9. 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>
10. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol* 1962:227–245. https://doi.org/10.1007/978-1-4939-9173-0_14
11. Love J, Palmer J, Stajich J, Esser T, Kastman E, Bogema D, Winter D. 2020. Funannotate: eukaryotic genome annotation pipeline.
12. 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>