

Draft genome sequence of *Ovulinia azaleae*, a quarantine pathogen known for causing azalea petal blight

Yunxin Zhang,¹ Yunfang Chen,² Sizhu Zheng,³ Xiaojun Yang,³ Zhenchuan Mao,¹ Guohua Chen,¹ Jian Ling¹

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We report the first draft genome sequence of *Ovulinia azaleae* (strain CBS 680.88). The long-read assembly resulted in a 56.83 Mb genome with an N50 of 3.05 Mb and 10,707 predicted protein-coding genes. The high-quality *Ovulinia azaleae* genome resource provides insights into the mechanisms of azalea petal blight.

KEYWORDS *Ovulinia azaleae*, azalea petal blight, genome

Ovulinia azaleae is a quarantine pathogen that exclusively infects azalea flowers (1, 2). Initial symptoms manifest as 1 mm diameter, water-soaked beige spots that progress to brown blotches, leading to flower wilting and premature abscission (3). This pathogen poses a significant threat to the rhododendron industry in China. To further analyze the pathogenic mechanism of *Ovulinia azaleae*, the whole genome sequence of *Ovulinia azaleae* was analyzed.

Strain CBS 680.88 was acquired from the Westerdijk Fungal Biodiversity Institute (CBS culture collection) as a pre-cultured and ready-to-use agar plate. For DNA extraction, fungal mycelia were grown in corn glucose broth at 25°C for 3 weeks, then harvested, washed, and freeze-dried. High-molecular weight genomic DNA was isolated using the CTAB method (4), followed by purification with the Monarch Genomic DNA Purification Kit (T3010L). Species identity was confirmed by sequencing the ITS region using primers ITS1/ITS4 (5). A BLASTn (v2.17.0+) search against the NCBI GenBank database (accessed 24 June 2025) identified *Ovulinia azaleae* (accession no. [Z73797.1](#)) as the top hit.

For multi-platform sequencing, both short-read and long-read libraries were constructed from the same genomic DNA extraction. An RNA-seq library was prepared using total RNA isolated from fungal mycelia. Quality control for the short-read and RNA-seq data was performed with fastp (v0.21.0) (6), which removed reads with mean base quality scores below Q20. For the long-read data, basecalling and adapter trimming were performed using Dorado (v0.9.0) with the "sup" model (dna_r10.4.1_e8.2_400bps_sup@v5.0.0). Reads with a mean Q-score below 10 were filtered out, yielding 15.04 Gb of high-quality long-read data. Detailed information regarding library preparation kits, sequencing platforms, quality control, and sequencing statistics are provided in Table 1.

Long reads were assembled independently using Canu (v2.2) (7) and NextDenovo (v2.5) (8), and the resulting assemblies were merged using QuickMerge (v0.3). The genome was polished via three rounds of Racon (v1.5.0) (9) with long reads, followed by three rounds of Pilon (v1.24) (10) with short reads. The final assembly genome was 56.83 Mb with 35.3% GC content, comprising 29 contigs with an N50 of 3.05 Mb. The largest contig spans 4.71 Mb (Fig. 1). For repetitive sequence annotation, a *de novo* library was constructed by RepeatModeler, and final repeats were masked by RepeatMasker (v4.0.9) (11), totaling 18.8 Mb (33.09% of the genome). The gene annotation for *Ovulinia azaleae* was generated by integrating *ab initio* and RNA-seq predictions. *Ab initio* predictions were completed through utilization of SNAP (v2.1.1) (12), AUGUSTUS

Editor Leighton Pritchard, University of Strathclyde, Glasgow, United Kingdom

Address correspondence to Jian Ling, lingjian@caas.cn, or Guohua Chen, chengguohua01@caas.cn.

Yunxin Zhang and Yunfang Chen contributed equally to this article. Author order was determined by drawing straws.

The authors declare no conflict of interest.

Received 29 July 2025

Accepted 7 November 2025

Published 15 January 2026

Copyright © 2026 Zhang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

TABLE 1 Sequencing data output statistics of *Ovulinia azaleae*^a

Data type	Short-read data	RNA-seq data	Long-read data
Library prep kit	Hieff NGS OnePot Pro DNA Library Prep Kit V4 (Yeasen Biotechnology)	VAHTS Universal V6 RNA-seq Library Prep Kit for MGI (Vazyme Biotech)	SQK-LSK1 14 kit (Oxford Nanopore Technologies) ^b
Platform	DNBSEQ-T7	DNBSEQ-T7	PromethION platform using R10.4.1 flow cells
PE	150 bp	150 bp	– ^d
Raw reads	102,425,560	45,185,756	2,332,836
Raw bases	15.36 Gb	6.78 Gb	15.41 Gb
Fold coverage	189×	115×	271×
Clean reads	102,418,162	45,181,656	2,277,562
Clean bases	15.29 Gb	6.69 Gb	15.04 Gb
N50	– ^d	– ^d	12,497
Quality control	fastp (v0.21.0, -q 20)	fastp (v0.21.0, -q 20)	Sup ^c (Q-score > 10)

^aLibrary preparation and sequencing were performed according to the manufacturer's instructions.

^bWithout fragmentation and size selection.

^cSup(dna_r10.4.1_e8.2_400bps_sup@v5.0.0).

^d–, not applicable.

(v3.3.2) (13), Gene Mark (v4.81), and Glimmer HMM (v3.0.4) (14). For transcriptome-based prediction, the transcripts were reconstructed with Spring Tie (v1.2.3) (15). These were integrated with EvidenceModeler (v1.1.1) (16), yielding 10,707 high-confidence genes. The assembly showed 97.2% completeness via BUSCO (v5.3.2) (17) against the fungi_odb10 dataset. In addition, functional annotation using NR, Swiss-Prot (2024_03), InterProScan (v88.0) (18), KOG, and Pfam (v35.0) (19) annotated 9,646 genes (90.09% of the total). Default parameters were used for all software unless otherwise specified.

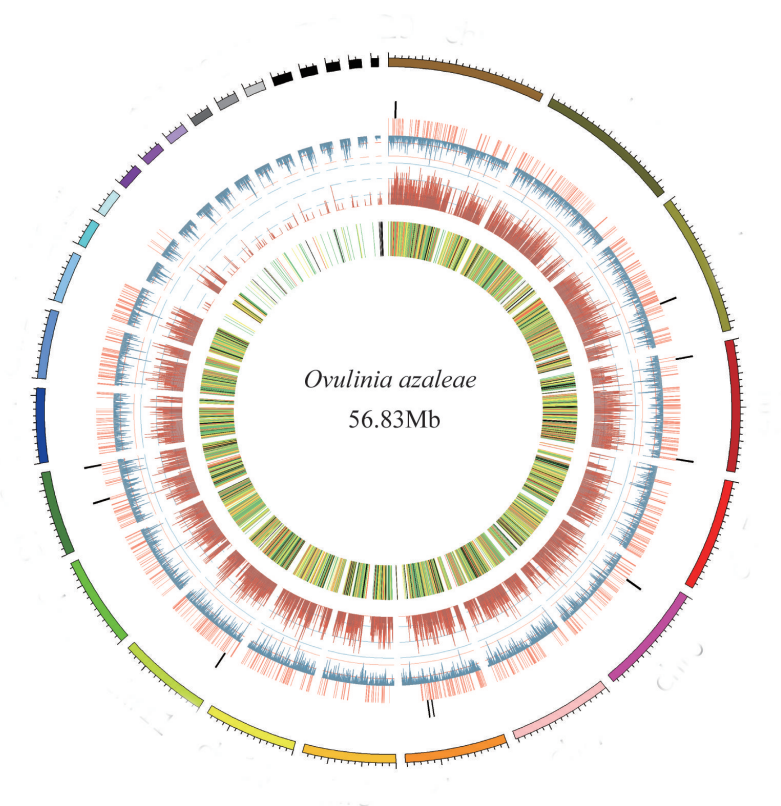


FIG 1 Genome features of *Ovulinia azaleae*. The outermost circle is the contigs. The bar charts from outside to inside in turn are secondary metabolite gene clusters (black), secreted proteins (orange), density of repetitive sequence (blue), gene density (dark red), and transcriptome (color heatmap).

ACKNOWLEDGMENTS

This research is funded by the National Key R&D Program of China (2023YFC2604900) and China Agriculture Research System (CARS-23).

AUTHOR AFFILIATIONS

¹State Key Laboratory of Vegetable Biobreeding, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China

²Comprehensive Technology Center of Suzhou Customs, Suzhou, China

³Animal, Plant and Food Inspection Center, Nanjing Customs, Nanjing, China

AUTHOR ORCIDs

Yunxin Zhang  <http://orcid.org/0009-0007-7433-5101>

Jian Ling  <http://orcid.org/0000-0001-5631-6950>

DATA AVAILABILITY

The genome and sequencing data have been deposited at NCBI under BioProject accession [PRJNA1282300](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1282300). The BioSample accession numbers for short-read sequencing, RNA-seq sequencing, and Nanopore sequencing are [SAMN49645637](https://www.ncbi.nlm.nih.gov/biosample/SAMN49645637), [SAMN49645639](https://www.ncbi.nlm.nih.gov/biosample/SAMN49645639), and [SAMN49645638](https://www.ncbi.nlm.nih.gov/biosample/SAMN49645638), respectively. The short-read sequencing reads are available in the NCBI Sequence Read Archive under accession number [SRR34207007](https://www.ncbi.nlm.nih.gov/sra/SRR34207007), the RNA-seq sequencing reads are accessible under accession number [SRR34207008](https://www.ncbi.nlm.nih.gov/sra/SRR34207008), and the Nanopore sequencing reads are accessible under accession number [ERR15807457](https://www.ncbi.nlm.nih.gov/sra/ERR15807457). This Whole Genome Shotgun project has been deposited in DDBJ/ENA/GenBank under accession number [JBPJSN000000000](https://www.ncbi.nlm.nih.gov/nuccore/JBPJSN000000000). The version described in this paper is the first version, [JBPJSN000000000](https://www.ncbi.nlm.nih.gov/nuccore/JBPJSN000000000). The genome annotation files generated in this study have been deposited in the Zenodo repository DOI: <https://doi.org/10.5281/zenodo.17311068>.

REFERENCES

- Lijun D, Huili Z, Xuellian L, Xianfeng C, Weijun D. 2020. Development of TaqMan MGB fluorescent real-time PCR assay for the detection of *Ovulinia azaleae*. *Acta Horticulturae Sinica* 47:797–804. <https://doi.org/10.16420/j.issn.0513-353x.2019-0465>
- Backhaus GF. 1994. Azalea-flower blight, caused by *Ovulinia azaleae* Weiss. *Acta Horti* 364:167–168. <https://doi.org/10.17660/ActaHortic.1994.364.22>
- Raabe R, Sciaroni R. 1955. Petal blight disease of azaleas: control of fungus-caused disease of azaleas and closely related plants is essential to prevent its further spread. *Calif Agric (Berkeley)* 9:7–14.
- Kuo P, Henderson IR, Lambing C. 2022. CTAB DNA extraction and genotyping-by-sequencing to map meiotic crossovers in plants. *Methods Mol Biol* 2484:43–53. https://doi.org/10.1007/978-1-0716-2253-7_4
- White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p 315–322. In *PCR Protocols: A Guide to Methods and Applications*.
- Shifu C, Yanqing Z, Yaru C, Jia G. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Jiang H, Zhou W, Zongyi S, Benxia H, Ayoola AO, Fan L, Jingjing L, Sandoval JR, Cooper DN, Kai Y, Jue R, Chuan-Le X, Depeng W, Dong-Dong W, Sheng W. 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol* 25:107. <https://doi.org/10.1186/s13059-024-03252-4>
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Qiangdong Z, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* Chapter 4:4. <https://doi.org/10.1002/0471250953.bi0410s25>
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>
- Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 Suppl 2:ii215–25. <https://doi.org/10.1093/bioinformatics/btg1080>
- 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>
- Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33:290–295. <https://doi.org/10.1038/nbt.3122>
- 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>
- Seppely 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
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res* 49:D412–D419. <https://doi.org/10.1093/nar/gkaa913>