

Genome sequence of *Pectobacterium brasiliense* WZZ2024, isolated from fast-growing cabbage in China

Zhen Zeng,^{1,2,3} Xiaorong Huang,^{2,3,4} Yin Jiang,^{2,3,4} Defeng Chen,^{1,3} Zunzheng Wei,^{2,3} Hao Liang^{1,3}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT The complete genome of *Pectobacterium brasiliense* WZZ2024, isolated from soft rot-infected petioles and tubers of fast-growing cabbage (*Brassica rapa* var. *glabra* Regel), comprises a single 4.83 Mb chromosome (GC content: 52.16%).

KEYWORDS genome

Pectobacterium brasiliense, a widely distributed bacterial phytopathogen, induces soft rot disease in economically significant crops (e.g., potato, kiwifruit, and cabbage) by producing pectolytic enzymes that degrade plant cell walls (1). This species exhibits heightened pathogenicity compared to other *Pectobacterium* spp. (2), causing tissue maceration and substantial agricultural losses.

P. brasiliense WZZ2024 was isolated from soft rot-infected petioles and tubers of fast-growing cabbage, collected in the Vegetable Research Center of Beijing Academy of Agriculture and Forestry Sciences, China (August 2022). Tissue samples were surface-sterilized in 75% ethanol, rinsed with sterile distilled water (SDW), and homogenized aseptically. The resulting homogenates were plated onto Luria-Bertani Agar (LB agar) and incubated at 28°C for 12 h. Single colonies were isolated by successive streaking to achieve purity. Axenic cultures derived from these colonies were grown, and genomic DNA was extracted using the FastPure Bacteria DNA Isolation Mini Kit (Vazyme Biotech, China) according to the manufacturer's instructions. The extracted DNA was subsequently subjected to whole-genome sequencing (3).

Library preparation and genome sequencing were performed at Biomarker Technologies (Beijing, China) on the PacBio Sequel II platform. Genomic DNA was sheared using g-TUBE (Covaris). The fragments were subjected to damage and end-repair using the SMRTbell Express Template Prep Kit 3.0 (PacBio, USA), followed by ligation of SMRTbell adapters and exonuclease digestion to remove unligated products. Size selection was performed using BluePippin (Sage Science, USA) to enrich fragments. Sequencing was conducted on the PacBio Sequel II platform using Sequel II Binding Kit 2.0 (PacBio, #101-842-900) and Sequel II Sequencing 2.0 Bundle (PacBio, #101-849-000) with HiFi chemistry v2.0. Post-processing steps retained high-quality CCS reads (>2 kb) using Filtlong v0.2.0 (--min_mean_q 20) for quality control. A total of 47,337 clean reads were obtained, with an average length of 9,475 bp. For Illumina sequencing, libraries were constructed and sequenced on the NovaSeq 6000 platform, yielding 3,292,417 read pairs, corresponding to 987.7 Mb of sequence data. Quality control was performed with Fastp v0.23.2 to remove adapters and low-quality reads. The Illumina data are available under SRA accession [SRR33800100](https://www.ncbi.nlm.nih.gov/sra/SRR33800100).

De novo assembly was performed using Unicycler v0.4.7 with both Illumina short reads and PacBio HiFi reads as input (4). Unicycler utilizes a multi-step hybrid approach: initial assembly graphs were constructed from the high-accuracy Illumina data using SPAdes v3.15.5, followed by bridging of contigs and gap-filling using the long PacBio reads. The assembly was subsequently polished with Pilon v1.22 (--mindepth 1 --changes

Editor Vanja Klepac-Ceraj, Wellesley College, Wellesley, Massachusetts, USA

Address correspondence to Hao Liang, Lianghao@nrcv.org.

Zhen Zeng and Xiaorong Huang contributed equally to this article. Author order was determined in order of increasing seniority.

The authors declare no conflict of interest.

Received 11 July 2025

Accepted 25 January 2026

Published 30 March 2026

Copyright © 2026 Zeng 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 Characteristics of *Pectobacterium brasiliense* WZZ2024

Characteristic	Result
Sequencing	
Raw reads	
No. of reads	47,456
Size (bp)	448,776,619
Clean reads	
No. of reads	47,337
Size (bp)	448,564,314
N ₅₀ (bp)	10,415
Average read length (bp)	9,475
Genome	
Assembly size (bp)	4,831,335
Contig	1
GC content (%)	52.16
No. of coding genes	4,212
Total gene size (bp)	4,178,577
Average gene length (bp)	992
Total repetitive sequence length (bp)	6,703
Repetitive sequence content (%)	0.14
rRNA	8, 7, 7 (5S, 16S, 23S)
tRNA	78
CRISPR	4
Genomic island	4
Prophage	4
Gene cluster	7
Protein subcellular localization	
Signal peptide	411
Transmembrane protein	1,038
Secreted protein	411

--fix all) using available Illumina short-read data for error correction. Circularization and orientation were performed using Circlator v1.5.5 (minimus2 --no_pre_merge). Default parameters were used for all software unless explicitly stated. The final genome assembly of *P. brasiliense* WZZ2024 consists of a single circular chromosome of 4,831,335 bp with a G+C content of 52.16% (Table 1). The topology of the contig was predicted to be circular.

Public genome assemblies were annotated using NCBI PGAP v5.3 (5), with coding sequences predicted via Prodigal v2.6.3 (6). Repeat elements were identified by RepeatMasker v4.0.5 (7), while tRNA and rRNA genes were annotated using tRNAscan-SE v2.0 and Infernal v1.1.3, respectively (8–10). Functional features were predicted through: CRISPR arrays: CRT v1.2 (11); Genomic islands: IslandPath-DIMOB v0.2 (12); Prophages: PhiSpy v2.3 (13); Biosynthetic clusters: antiSMASH v5.0.0 (14). Secretory protein candidates were identified using SignalP after TMHMM-based transmembrane domain filtering (15, 16). Analysis results are listed in Table 1.

ACKNOWLEDGMENTS

This work was supported by the Scientific and technological innovation capacity building project of Beijing Academy of Agriculture and Forestry Sciences (KJCX20251312, CYJS202502).

AUTHOR AFFILIATIONS

¹National Engineering Research Center for Vegetables, Beijing Vegetable Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

²Institute of Grassland, Flowers and Ecology, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

³Key Laboratory of Urban Agriculture (North China), Ministry of Agriculture and Rural Affairs, Beijing, China

⁴College of Horticultural Science & Technology, Hebei Key Laboratory of Horticultural Germplasm Excavation and Innovative Utilization/Hebei Higher Institute Application Technology Research and Development Center of Horticultural Plant Biological Breeding, Hebei Normal University of Science and Technology, Qinhuangdao, China

AUTHOR ORCIDS

Zhen Zeng  <http://orcid.org/0009-0003-6644-9855>

Hao Liang  <http://orcid.org/0000-0003-1253-6331>

AUTHOR CONTRIBUTIONS

Zhen Zeng, Conceptualization, Writing – original draft, Writing – review and editing | Xiaorong Huang, Writing – original draft, Writing – review and editing | Yin Jiang, Data curation | Defeng Chen, Writing – review and editing | Zunzheng Wei, Conceptualization | Hao Liang, Writing – review and editing

DATA AVAILABILITY

This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under the accession [CP195137.1](#) and BioProject accession number [PRJNA1271328](#). The raw reads have been submitted to the SRA under accession numbers [SRR33800099](#) and [SRR33800100](#). The version described in this paper is the first version.

REFERENCES

- Meng X, Chai A, Shi Y, Xie X, Ma Z, Li B. 2017. Emergence of bacterial soft rot in cucumber caused by *Pectobacterium carotovorum* subsp. *brasiliense* in China. *Plant Dis* 101:279–287. <https://doi.org/10.1094/PDIS-05-16-0763-RE>
- Oulghazi S, Sarfraz S, Zaczek-Moczydlowska MA, Khayri S, Ed-Dra A, Lekbach Y, Campbell K, Novungayo Moleleki L, O'Hanlon R, Faure D. 2021. *Pectobacterium brasiliense*: genomics, host range and disease management. *Microorganisms* 9:106. <https://doi.org/10.3390/microorganisms9010106>
- Jiang Y, Zeng Z, Wang Z, Wang Y, Gou R, Huang X, Lu Y, Zhang G, Wei Z. 2025. Genome sequence of *Providencia rettgeri* ZAPR22R, isolated from calla lily in China. *Microbiol Resour Announc* 14. <https://doi.org/10.1128/mra.01049-24>
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
- 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>
- Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res* 46:D335–D342. <https://doi.org/10.1093/nar/gkx1038>
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29:2933–2935. <https://doi.org/10.1093/bioinformatics/btt509>
- Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences, p 1–14. *In* Kollmar M (ed), *Gene prediction: methods and protocols*. Springer, New York, NY.
- Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. 2007. CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. <https://doi.org/10.1186/1471-2105-8-209>
- Bertelli C, Brinkman FSL. 2018. Improved genomic island predictions with IslandPath-DIMOB. *Bioinformatics* 34:2161–2167. <https://doi.org/10.1093/bioinformatics/bty095>
- Akhter S, Aziz RK, Edwards RA. 2012. PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res* 40:e126. <https://doi.org/10.1093/nar/gks406>
- Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 47:W81–W87. <https://doi.org/10.1093/nar/gkz310>
- Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786. <https://doi.org/10.1038/nmeth.1701>
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315>