

# Draft genome of carotenoid-producing endophytic *Pseudomonas* sp. 102515 from *Taxus chinensis*

Ozkan Fidan<sup>1</sup>**AUTHOR AFFILIATION** See affiliation list on p. 2.

**ABSTRACT** Here, we report the draft genome sequence of endophytic *Pseudomonas* sp. 102515 isolated from *Taxus chinensis* collected from Logan, UT, USA. The genome is composed of 36 contigs and around 4.9 Mbp in size. The GC content is 66% with an  $N_{50}$  length of 918.9 kbp and  $L_{50}$  count of 2.

**KEYWORDS** genome, *Pseudomonas*, endophytic bacteria, *Taxus chinensis*

Endophytic microbes play significant roles in plant health by providing various benefits such as nitrogen fixation, phosphate solubilization, and growth promotion. Additionally, they are a rich source of bioactive natural products (1). Endophytic *Pseudomonas* sp. 102515 were isolated from the leaves of *Taxus chinensis* and is a high-yield producer of zeaxanthin diglucoside. This isolate can also produce turnerbactin, a siderophore involved in nitrogen fixation. Therefore, *Pseudomonas* sp. 102515 can have plant growth-promoting activity (2). The closest relative strain to this endophyte is *Pseudomonas psychrotolerans*, exhibiting plant growth-promoting activity (3).

The leaves of *T. chinensis* were collected on the Logan campus of Utah State University. The details of isolation and identification were provided in our previous article (2). After the disinfection process, the leaves were ground in an aqueous NaCl solution (0.85%) to release endophytic microorganisms. The tissue extract was diluted and streaked on lysogeny broth (LB) media for the isolation of bacteria. One of the colonies led to the pigment-producing strain, which was identified as *Pseudomonas* sp. 102515 (Taxonomy ID: 3071568).

The endophytic *Pseudomonas* sp. 102515 were grown in LB media from a single colony at 28°C. The genomic DNA was extracted with the EasyPure Genomic DNA Kit (Transgen Biotech). Subsequently, genomic DNA were sent to Refgen Biotech Inc. for sequencing. In the library preparation, the Illumina DNA Prep (Illumina, USA, #20060059) kit was used. During the library preparation, fragmentation, addition of index/barcode sequences, amplification, and purification were carried out by following the manufacturer's kit instructions. Then, the length distribution of the prepared library was measured using 2100 Bioanalyzer (Agilent, USA, #G2939BA). After quality verification, approximately 5 million paired-end reads with a length of 150 bp were obtained from NovaSeq 6000. 6.24 Gbp of DNA were sequenced, totaling 4,876,136 bp of DNA, which represents 1,279× coverage of the genome. After the sequencing process, the FASTQC (v.1.0.0) tool was used for the quality control of the raw reads. The Trimmomatic (v.0.39) tool was used for cropping operations according to quality values. The Shovill (v.1.1.0) tool and SPAdes assembler (v.3.9.0) were used in *de novo* assembly processes (4, 5).

The genome is composed of 36 contigs and around 4.9 Mbp in size. The GC content is 66% with an  $N_{50}$  length of 918.9 kbp and  $L_{50}$  count of 2. The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) was used for annotation (6). Based on the PGAP (v.4.9), the

**Editor** David A. Baltrus, The University of Arizona, Tucson, Arizona, USA

Address correspondence to Ozkan Fidan, ozkan.fidan@agu.edu.tr.

The author declares no conflict of interest.

**Received** 12 February 2024

**Accepted** 11 June 2024

**Published** 28 June 2024

Copyright © 2024 Fidan. 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/).

genome includes 4,575 genes and 4,444 protein-coding sequences. CheckM analysis (v.1.2.2) was used to calculate the completeness of the genome, which was found as 98.56%, while 0.61% contamination was reported (7). Comprehensive genome analysis was performed using the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) (v.3.35.5) (8). The annotation RAST tool kit (RASTtk) (v.1.073) had 900 hypothetical proteins and 3,615 proteins with functional assignments (9). The proteins with functional assignments included 1,136 proteins with Enzyme Commission numbers (10), 973 with Gene Ontology assignments (11), and 862 proteins that were mapped to KEGG pathways (12). According to BV-BRC annotation, this genome had 4,287 proteins that belong to the genus-specific protein families (PLFams) and 4,347 proteins that belong to the cross-genus protein families (PGFams) (13).

## ACKNOWLEDGMENTS

This study was supported by the Scientific and Technological Research Council of Türkiye with project number 221Z280.

## AUTHOR AFFILIATION

<sup>1</sup>Department of Bioengineering, Faculty of Life and Natural Sciences, Abdullah Gul University, Kayseri, Turkey

## AUTHOR ORCIDs

Ozkan Fidan  <http://orcid.org/0000-0001-5312-4742>

## AUTHOR CONTRIBUTIONS

Ozkan Fidan, Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Writing – original draft, Writing – review and editing

## DATA AVAILABILITY

The Illumina raw reads were uploaded to the SRA database under the accession number of [SRX24173677](https://www.ncbi.nlm.nih.gov/sra/SRX24173677). The generated contigs were uploaded on NCBI under Biosample ID [SAMN37118173](https://www.ncbi.nlm.nih.gov/biosample/SAMN37118173). Two contigs smaller than 200 bp were discarded during the Genbank upload. The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under accession [JAVHTY0100000000](https://www.ncbi.nlm.nih.gov/nuccore/JAVHTY0100000000). The version described in this paper is version [JAVHTY0100000000.1](https://www.ncbi.nlm.nih.gov/nuccore/JAVHTY0100000000).

## REFERENCES

1. Tripathi A, Pandey P, Tripathi SN, Kalra A. 2022. Perspectives and potential applications of endophytic microorganisms in cultivation of medicinal and aromatic plants. *Front Plant Sci* 13:985429. <https://doi.org/10.3389/fpls.2022.985429>
2. Fidan O, Zhan J. 2019. Discovery and engineering of an endophytic *Pseudomonas* strain from *Taxus chinensis* for efficient production of zeaxanthin diglucoside. *J Biol Eng* 13:66. <https://doi.org/10.1186/s13036-019-0196-x>
3. Liu R, Zhang Y, Chen P, Lin H, Ye G, Wang Z, Ge C, Zhu B, Ren D. 2017. Genomic and phenotypic analyses of *Pseudomonas psychrotolerans* PRS08-11306 reveal a turnerbactin biosynthesis gene cluster that contributes to nitrogen fixation. *J Biotechnol* 253:10–13. <https://doi.org/10.1016/j.jbiotec.2017.05.012>
4. 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>
5. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using spades *de novo* assembler. *Curr Protoc Bioinforma* 70:e102. <https://doi.org/10.1002/cpbi.102>
6. Tatusova T, DiCuccio M, Badretdin A, Chetvermin 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>
7. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
8. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, Dempsey DM, Dickerman A, Dietrich EM, Kenyon RW, et al. 2023. Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res* 51:D678–D689. <https://doi.org/10.1093/nar/gkac1003>
9. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R,

- Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>
10. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D. 2004. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* 32:D431–3. <https://doi.org/10.1093/nar/gkh081>
  11. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene Ontology: tool for the unification of biology. *Nat Genet* 25:25–29. <https://doi.org/10.1038/75556>
  12. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–62. <https://doi.org/10.1093/nar/gkv1070>
  13. Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, Shukla M, Vonstein V, Wattam AR, Yoo H. 2016. PATtyFams: protein families for the microbial genomes in the PATRIC database. *Front Microbiol* 7:118. <https://doi.org/10.3389/fmicb.2016.00118>

GCRLS