BMC Genomic Data

DATA NOTE Open Access

Whole-genome sequence of *Pseudomonas* sp. strain HOU2 isolated from dangshen (*Codonopsis javanica*) roots

Van Hong Thi Dao[1], Loan To Nguyen[2], Khanh Phuong Do[1], Vinh The Nguyen[1], Hieu Van Nguyen[3], Khanh Ngoc Pham[4], Truong Xuan Nguyen[5] and Son Truong Dinh[5,6*]

Abstract

Objectives This study aims to generate a *de novo* complete whole-genome assembly of *Pseudomonas* sp. strain HOU2, which is an endophytic bacterium isolated from dangshen roots that shows to improve the growth of in vitro dangshen plants. Further investigation of the whole genome of *Pseudomonas* sp. strain HOU2 will help identify potential genes or pathways that could be involved in the plant growth-promoting effects on in vitro dangshen plants, providing valuable information for future applications.

Data description The genomic DNA of *Pseudomonas* sp. strain HOU2 was sequenced using Oxford Nanopore's PromethION sequencer with an R10.4.1 flow cell (Table 1, Data file 1). The assembly of the *Pseudomonas* sp. strain HOU2 genome was conducted using Flye version 2.9, resulting in a single circular chromosome of 6,047,544 bp with a mean coverage of 488 (Table 1, Data file 2). The annotation of genes, proteins, and features of the HOU2 genome were performed by the RAST server (Rapid Annotation using Subsystem Technology) (https://rast.nmpdr.org/) (Table 1, Data file 3, 4, 5) [6, 7]. The *Pseudomonas* sp. strain HOU2 genome was determined to be most similar to that of *Pseudomonas koreensis* using the Type Strain Genome Server (https://tygs.dsmz.de/, version v391) [8].

Keywords Endophytic bacterium, PromethION, *Pseudomonas* sp. strain HOU2, Whole-genome sequencing

Objective

The study of plant growth-promoting bacteria can provide important information for the application of beneficial bacteria in sustainable crop production. Endophytic bacterial inoculation increased stalk length, root number, fresh mass and dry mass [1]. Endophytic bacteria play crucial roles in the health and development of medicinal plants. They promote plant growth and increase resistance against pathogens and environmental stresses. Moreover, endophytes also regulate the synthesis of secondary metabolites of host plants or produce bioactive metabolites on their own without interacting with the host [2–4].

*Correspondence:
Son Truong Dinh
dtson@vnua.edu.vn
[1]Hanoi Open University, B101 Nguyen Hien Street, Hai Ba Trung District, Hanoi, Vietnam
[2]Centre for Animal Science, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, Qld 4072, Australia
[3]Institute of Biotechnology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Street, Cau Giay District, Hanoi, Vietnam
[4]National Institute of Medicinal Materials, Hanoi, Vietnam
[5]Institute of Agrobiology, Vietnam National University of Agriculture, Hanoi, Vietnam
[6]Faculty of Biotechnology, Vietnam National University of Agriculture, Hanoi, Vietnam

Table 1 Overview of data files/ sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	Genome sequence data of *Pseudomonas* sp. HOU2	FASTQ file (.fastq)	NCBI SRA Database Accession number SRR29666724 https://identifiers.org/ncbi/insdc.sra:SRR29666724 [15]
Data file 2	*Pseudomonas* sp. HOU2 chromosome, complete genome	Fasta files nucleic acids (.fasta)	NCBI GenBank Accession number CP160398 https://identifiers.org/ncbi/insdc:CP160398 [16]
Data file 3	*Pseudomonas* sp. HOU2 predicted gene sequences	Fasta files nucleic acids (.fasta)	Figshare https://doi.org/10.6084/m9.figshare.26325310) [17]
Data file 4	*Pseudomonas* sp. HOU2 predicted protein sequences	Fasta files amino acids (.fasta)	Figshare https://doi.org/10.6084/m9.figshare.26325340) [18]
Data file 5	*Pseudomonas* sp. HOU2 spreadsheet of predicted features	Excel file (.xls)	Figshare: https://doi.org/10.6084/m9.figshare.26325049) [19]

Dangshen (*Codonopsis javanica*) has been used as a medicinal herbal plant to treat diabetes and other diseases [5]. In Vietnam, dangshen has been cultivated in many provinces, such as Lai Chau, Dien Bien, and Lao Cai. The use of beneficial microorganisms in the production of dangshen is suitable for sustainable medicinal plant value chain development. Therefore, the isolation, characterization and study of the whole genome of bacterial endophytes may help improve the yield and quality of medicinal plants.

We isolated and characterized the plant growth-promoting effects of several endophytic bacteria isolated from *C. javanica* roots, of which the *Pseudomonas* sp. strain HOU2 improved the growth of *in vitro C. javanica* plants and was selected for whole-genome sequencing investigation. Here, we present the complete whole genome and gene annotation predicted by the RAST server, [6–8] for *Pseudomonas* sp. strain HOU2, which could serve as a resource and dataset for investigating genes and pathways associated with plant growth-promoting effects.

Data description

Roots from healthy *Codonopsis javanica* plants were subjected to surface sterilization with 0.5% NaOCl for 20 min. One hundred microliters of the third rinse water were placed on Luria broth (LB) medium and incubated at 27 °C for 3 days to check for surviving colonies. The surface-disinfected roots were smashed in a sterilized Petri dish, diluted in liquid LB, and then spread on LB media. Pure culture isolation was achieved by re-streaking twice on LB agar medium [1].

A single colony of the HOU2 bacteria was grown overnight in liquid LB media in a shaker (27 °C, 200 rpm). Genomic DNA extraction was conducted using a modified protocol [9]. The integrity of the HOU2 genomic DNA was verified on 1% agarose gel electrophoresis. DNA quantity and quality were evaluated using the Qubit dsDNA Broad Range assay kit (Thermo Fisher Scientific) and NanoDropND 1000 (v.3.5.2, Thermo Fisher Scientific). The genomic DNA was sequenced with an Oxford Nanopore's PromethION sequencer using R10.4.1 flow cell.

The extracted DNA sample was prepared using the Native Barcoding Kit 24 V14 protocol (SQK-NBD114.24) from Oxford Nanopore Technologies (Oxford, United Kingdom). The library was prepared according to the manufacturer's instructions, with the exception of increasing the incubation time to 30 min for both the end-repair and ligation steps. To avoid or reduce DNA shearing, bore tips and gentle sample flicking (without vortexing) were used. Raw signal files (FAST5) were used for base calling with Guppy v6.4.6 in super accuracy mode, resulting in over 3.9 million reads (Q score ≥ 10) with an average read length of 7.4 kb. Reads under 5000 bp and with a Q score ≤ 20 were removed by Chopper version 0.5, resulting in over 262 thousand reads with a read length N50 of 12.8 kb [10].

The pre-processed reads were *de novo* assembled using Flye v. 2.9-b1768 with the parameters --nano-hq and --read-error 0.03 for the Nanopore super accuracy base call file [11] to obtain the complete whole genome of *Pseudomonas* sp. strain HOU2. The quality and completeness of the assembly were assessed using BUSCO v5.4.5 [12] in prok_genome mode (the lineage dataset is bacteria_odb10, creation date: 2020-03-06, number of genomes: 4085, number of BUSCOs: 124), which indicated that the whole genome of the HOU2 sequence is > 99.2% complete (Table 1, Data set 2). In addition, the genome was also evaluated by NCBI, using the Prokaryotic Genome Annotation Pipeline (PGAP) gene set with the *Pseudomonas* CheckM marker set (v1.2.3) [13], resulting in a completeness of 99.68% (100th percentile) and contamination of 0.08% (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_040729435.1/).

The HOU2 genome was most similar to that of *Pseudomonas koreensis* strain FP1691, accession number: GCA_026314355.1, with an original ANI score of 93.68 and an OrthoANI score of 94.01%, as determined by CJ Bioscience's Orthologous Average Nucleotide Identity Tool (OAT) [8, 14].

Limitations

Here, we used Nanopore long-read sequencing to produce a complete, high-quality single circular genome of *Pseudomonas* sp. HOU2. The main limitation of this data note is the genome assembly completeness of the HOU2 strain was >99.2%, which was not yet fully covered. In addition, this data note was limited to the description of how the datasets were generated. Further data analysis of the biosynthesis-related gene clusters associated with plant growth-promoting effects should be performed to explain why and how *Pseudomonas* sp. HOU2 improves *Codonopsis javanica* growth performance.

Abbreviations

LB	Luria broth
DNA	Deoxyribonucleic acid
BUSCO	Benchmarking Universal Single-Copy Ortholog
IAA	Indole-3-acetic acid

Acknowledgements

We would like to thank the Hanoi Open University for the support and funding this project (MHN2022-01.26) and the NCBI staffs (https://www.ncbi.nlm.nih.gov/), who did the annotation task for this genome.

Author contributions

Conceptualization, STD, LTN, VHTD; investigation, STD, LTN, KPD, VTN, HVN, KNP, TXN; resources, VHTD; writing, STD, LTN, VHTD; supervision, STD, LTN, VHTD. The authors read and approved the final manuscript.

Funding

The authors thank Hanoi Open University for supporting this study (MHN2022-01.26).

Data availability

The data described in this Data Note can be freely and openly accessed at NCBI or Figshare database, including: Data file 1: Genome sequence data of Pseudomonas sp. HOU2, FASTQ file (.fastq), NCBI Sequence Read Archive, accession number: SRR29666724 (https://www.ncbi.nlm.nih.gov/sra/SRR29666724).Data file 2: Pseudomonas sp. HOU2 chromosome, complete genome, Fasta files nucleic acids (.fasta), NCBI accession number: CP160398.1 (https://www.ncbi.nlm.nih.gov/nuccore/CP160398).Data file 3: Pseudomonas sp. HOU2 predicted gene sequences, Fasta files nucleic acids (.fasta), Figshare (https://doi.org/10.6084/m9.figshare.26325310).Data file 4: Pseudomonas sp. HOU2 predicted protein sequences, Fasta files amino acids (.fasta), Figshare (https://doi.org/10.6084/m9.figshare.26325340).Data file 5: Pseudomonas sp. HOU2 spreadsheet of predicted features, Excel file (.xls), Figshare: (https://doi.org/10.6084/m9.figshare.26325049).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 22 July 2024 / Accepted: 12 December 2024
Published online: 26 December 2024

References

1. Dao VHT, Do KP, Nguyen HV, Nguyen CM, Thi Tran DT, Thanh Dang TT, Nguyen TX, Dinh ST. Identification and evaluation of the growth promotion of endophytic bacteria on *in vitro* potato plants. Pakistan J Biol sciences: PJBS. 2023;26(7):371–9.
2. Ek-Ramos MJ, Gomez-Flores R, Orozco-Flores AA, Rodriguez-Padilla C, Gonzalez-Ochoa G, Tamez-Guerra P. Bioactive products from plant-endophytic gram-positive bacteria. Front Microbiol. 2019;10:463.
3. Mostafa F. The endophytes: A new resource for vulnerable plant bioactive compounds. In: *Medicinal Plants.* Edited by Mozaniel Santana de O, Eloisa Helena de Aguiar A, Ravendra K, Suraj NM. Rijeka: IntechOpen; 2024: Ch. 3.
4. Lodi RS, Dong X, Jiang C, Sun Z, Deng P, Sun S, Wang X, Wang H, Mesa A, Huang X et al. Antimicrobial activity and enzymatic analysis of endophytes isolated from *Codonopsis pilosula*. FEMS Microbiol Ecol 2023, 99(8).
5. Ueda JY, Tezuka Y, Banskota AH, Le Tran Q, Tran QK, Harimaya Y, Saiki I, Kadota S. Antiproliferative activity of Vietnamese medicinal plants. Biol Pharm Bull. 2002;25(6):753–60.
6. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.
7. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 2015;5:8365.
8. Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 2019;10(1):2182.
9. Masoomi-Aladizgeh F, Jabbari L, Nekouei RK, Aalami A. A simple and rapid system for DNA and RNA isolation from diverse plants using handmade kit. Protoc Exch 2016.
10. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinf (Oxford England) 2023, 39(5).
11. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6.
12. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinf (Oxford England). 2015;31(19):3210–2.
13. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.
14. Lee I, Ouk Kim Y, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016;66(2):1100–3.
15. Dao VHT, Dinh ST. Data file 1. NCBI Sequence Read Archive of *Pseudomonas* sp. HOU2, https://identifiers.org/ncbi/insdc.sra:SRR29666724 2024.
16. Dao VHT, Dinh ST. Data file 2. Complete genome of *Pseudomonas* sp. HOU2 on NCBI, https://identifiers.org/ncbi/insdc:CP160398. 2024.
17. Dao VHT, Dinh ST. Data file 3. *Pseudomonas* sp. HOU2 predicted gene sequences, https://doi.org/10.6084/m9.figshare.26325310. 2024.
18. Dao VHT, Dinh ST. Data file 4. *Pseudomonas* sp. HOU2 predicted protein sequences, https://doi.org/10.6084/m9.figshare.26325340. 2024.
19. Dao VHT, Dinh ST. Data file 5. *Pseudomonas* sp. HOU2 spreadsheet of predicted features https://doi.org/10.6084/m9.figshare.26325049. 2024.

Publisher's note