Data Article

Data on complete genome sequence and annotation of *Paenibacillus sonchi* LMG 24727<sup>T</sup>

GyuDae Lee, Min-Ji Kim, Jae-Ho Shin<sup>∗</sup>

Department of Applied Biosciences, Kyungpook National University, Daegu 41566, Republic of Korea

**Article info**

Received 25 May 2021
Revised 27 June 2021
Accepted 3 August 2021
Available online 8 August 2021

**Keywords:**
Complete genome
Nanopore technology
Nitrogen metabolism
*Paenibacillus sonchi* LMG 24727

**Abstract**

*Paenibacillus sonchi* LMG 24727<sup>T</sup> was acquired from the Belgian Coordinated Collections of Microorganisms (BCCM) isolated from Sonchus oleraceus rhizosphere soil. This strain is a gram-positive, aerobic, rod-shaped bacterium. The strain's genomic DNA was extracted using a Wizard<sup>®</sup> Genomic DNA Purification Kit, and whole-genome sequencing was performed using the Nanopore MinION platform. Whole-genome assembly was performed using Flye assembler, and a total 7,782,254 bp length of circular chromosome and two plasmids was assembled by using a 1,558,445,868 bp length of raw reads. Genome annotation by the Prokaryotic Genome Annotation Pipeline (PGAP) showed the complete genome to contain 50.6% G+C content; 6264 protein-coding genes; 27 rRNA genes; 89 tRNA genes; and 4 ncRNA genes. Additionally, multiple genes related to nitrogen metabolism were annotated from the Rapid Annotation using Subsystem Technology (RAST) server. The complete genome sequence data have been submitted to the National Center for Biotechnology Information (NCBI) and have been deposited at DDBJ/ENA/GenBank under the accession number CP068595.1, CP068596.1, and CP068597.1.

© 2021 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license ([http://creativecommons.org/licenses/by-nc-nd/4.0/](http://creativecommons.org/licenses/by-nc-nd/4.0/))

---

<sup>∗</sup> Corresponding author.

_E-mail address: jhshin@knu.ac.kr_ (J.-H. Shin).

https://doi.org/10.1016/j.dib.2021.107271
2352-3409/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license ([http://creativecommons.org/licenses/by-nc-nd/4.0/](http://creativecommons.org/licenses/by-nc-nd/4.0/))
Specifications Table

| Subject       | Agricultural Microbiology |
|---------------|---------------------------|
| Specific subject area | Genomics and Bioinformatics |
| Type of data  | Complete genome sequence in FASTA format |
| How data were acquired | Genome sequencing was performed by Oxford Nanopore MinION platform; base calling by Guppy v4.4.1; genome assembly by Flye v. 2.8.2-b1691; genome annotation by National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline (PGAP) and Rapid Annotation using Subsystem Technology (RAST) |
| Data format   | Genomic DNA was extracted using Wizard® Genomic DNA Purification Kit (Promega, USA). |
| Parameters for data collection | The reads of Paenibacillus sonchi LMG 24727T produced from Nanopore MinION platform and de novo assembled into three contigs using Flye assembler. The genome annotation was completed using PGAP and RAST. |
| Description of data collection | Genome sequence data source: Kyungpook National University, Daegu, Republic of Korea |
| Data source location | Repository name: National Center for Biotechnology Information (NCBI) |
| Data accessibility | Direct URL to data: https://www.ncbi.nlm.nih.gov/assembly/GCA_016772475.1 |

Value of the Data

- The complete genome sequence data of *P. sonchi* LMG 24727T provide essential information and insight for bacterial nitrogen metabolism.
- The genome data of *P. sonchi* LMG 24727T accelerate knowledge for agricultural applications and in all microbial research communities.
- As *P. sonchi* LMG 24727T is a type strain, the genome sequence data are useful for comparative genomic studies.

1. Data Description

The type strain *P. sonchi* LMG 24727T having nitrogen fixing ability is a type strain, isolated from rhizosphere soil of *Sonchus oleraceus* in Xinjiang, China [1]. Although this strain has potential plant growth-promoting activity, the completed genome sequence has not yet been provided. Therefore, it is important to obtain a high-quality genome sequence for this strain. Whole-genome sequencing of *P. sonchi* LMG 24727T using Oxford Nanopore Technologies’ (ONT) MinION platform generated a total of 134,596 reads with 1,558,445,868 bp. The N50 of sequencing reads was 25,628 bp and the mode Phred quality score was 12.45. *De novo* assembly was performed using Flye assembler, and the circular completed chromosome and two plasmids were constructed in CGView (Fig. 1) [2]. The draft genome contains 7,782,254 bp with 50.6% of G+C content. The genome was annotated using NCBI PGAP [3] with the best-placed reference protein set GeneMarkS-2+; therefore, a total of 6264 protein-coding genes and 120 RNA genes (27 rRNA genes, 89 tRNA genes, and 4 ncRNA genes) were predicted (Table 1).

A complete genome annotation was conducted using RAST server [4], and a total of 7956 coding sequences was identified. In addition, a total of 354 subsystems was classified with 19% of subsystems coverage. Subsystem features belonged mostly to carbohydrates (452 genes); followed by amino acids and derivatives (304 genes); protein metabolism (241 genes); cofactors, vitamins, prothetic groups, and pigments (175 genes); and nucleosides and nucleotides (127
Fig. 1. Complete genome sequence of *Paenibacillus sonchi* LMG 24727<sup>T</sup>. The circular map of this strain’s genome sequence was constructed using CGView.

Table 1

| Genomic feature                  | Value       |
|----------------------------------|-------------|
| Genome length (bp)               | 7,782,254   |
| G+C content (%)                  | 50.6        |
| Total number of genes            | 7019        |
| Number of protein-coding genes   | 6264        |
| Total number of RNA genes        | 120         |
| rRNA genes (5S, 16S, 23S)        | 9, 9, 9     |
| tRNA genes                       | 89          |
| ncRNA genes                      | 4           |
| Pseudo genes                     | 635         |

Of note, this strain had 37 genes corresponding to the nitrogen metabolism subsystem, including nitrosative stress (9 genes), nitrate and nitrite ammonification (12 genes), ammonia assimilation (10 genes), and denitrifying reductase gene clusters (6 genes) (Fig. 2 and Supplementary data 1).

2. Experimental Design, Materials and Methods

*P. sonchi* LMG4727<sup>T</sup> was obtained from the BCCM/LMG bacteria collection ([http://www.belspo.be/bccm/](http://www.belspo.be/bccm/)) and grown on TSA medium at 30 °C for 24 h at 200 rpm. After overnight
culture, genomic DNA was extracted using a Wizard® Genomic DNA Purification Kit (Promega, USA). Unlike other sequencing platforms such as PacBio and Illumina, Oxford Nanopore technology did not require shearing the genomic DNA to a specific size. The quality and quantity of DNA were measured using a Nanodrop One Spectrophotometer (Thermo Fisher Scientific, USA) and Qubit 3.0 fluorometer (Thermo Fisher Scientific, USA), respectively.

A ligation sequencing kit (SQK-LSK110, ONT, UK) and NEBNext Companion Module for Oxford Nanopore Technologies Ligation Sequencing (NEB, USA) were used for DNA end-repairing, dA-tailing, and adapter ligation. The final library for MinION sequencing was loaded into the flow cell (R10.3, ONT, UK) and sequencing was performed for 20 h. FAST5 files were generated and base calling was processed using the Guppy v4.4.1 software package (Ubuntu 18 GPU, GeForce GTX 1660) [5]. Reads processed base calling with an average Phred quality score lower than 7 were discarded, and FASTQ files were generated.

De novo assembly with Nanopore reads was performed using Flye v. 2.8.2-b1691 (options: flye–nano-raw–genome-size 7.5 and default settings for remaining) [6]. The complete genome annotation was carried out using NCBI PGAP version 2021-01-11.build5132 [3] and the RAST server [4].

CRediT Author Statement

GyuDaee Lee: Conceptualization, Methodology, Visualization, Writing – original draft preparation; Min-Ji Kim: Data curation, Original draft preparation; Jae-Ho Shin: Conceptualization, Supervision, Writing – reviewing & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.
Acknowledgments

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Agricultural Microbiome R&D Program funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (918010044SB010).

Supplementary Materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.dib.2021.107271.

References

[1] Y.Y. Hong, Y.C. Ma, Y.G. Zhou, F. Gao, H.C. Liu, S.F. Chen, Paenibacillus sonchi sp. nov., a nitrogen-fixing species isolated from the rhizosphere of Sonchus oleraceus, Int. J. Syst. Evol. Microbiol. 59 (Pt 11) (2009) 2656–2661 Nov, doi:10.1099/ijs.0.009308-0.
[2] P. Stothard, D.S. Wishart, Circular genome visualization and exploration using CGView, Bioinformatics 21 (2005) 537–539, doi:10.1093/bioinformatics/bti054.
[3] T. Tatusova, M. DiCuccio, A. Badrettin, V. Chetvernin, E.P. Nawrocki, L. Zaslavsky, A. Lomsadze, K.D. Pruitt, M. Borodovsky, J. Ostell, NCBI prokaryotic genome annotation pipeline, Nucleic Acids Res. 44 (2016) 6614–6624, doi:10.1093/nar/gkw569.
[4] R.K. Aziz, D. Bartels, A.A. Best, M. Dejongh, T. Disz, R.A. Edwards, K. Formsmna, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, The RAST server: rapid annotations using subsystems technology, BMC Genom. 9 (2008) 1–15, doi:10.1186/1471-2164-9-75.
[5] R.R. Wick, L.M. Judd, K.E. Holt, Performance of neural network basecalling tools for Oxford nanopore sequencing, Genome Biol. 20 (2019) 1–10, doi:10.1186/s13059-019-1727-y.
[6] M. Kolmogorov, J. Yuan, Y. Lin, P.A. Pevzner, Assembly of long, error-prone reads using repeat graph, Nat. Biotechnol. 37 (2019) 540–546, doi:10.1038/s41587-019-0072-8.