

Draft genome sequence of *Paenibacillus* sp. M.A.Huq-81 isolated from the soil sample of a pine garden located in Anseong, South Korea

Md. Amdadul Huq,¹ Jong-Whi Park,¹ Md. Shahedur Rahman²

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT This study reports the draft genome sequence of *Paenibacillus* sp. M.A.Huq-81, a bacterial strain isolated from the soil sample of a pine garden located in Anseong, South Korea, in 2021. The genome consists of 5,210,425 base pairs assembled into 22 contigs, encoding 4,797 predicted protein-coding genes.

KEYWORDS draft genome sequence, *Paenibacillus* sp. M.A.Huq-81, soil sample of a pine garden

The genus *Paenibacillus*, first identified by Ash et al. (1), is classified under the family *Paenibacillaceae*. To date, it includes 323 validly published species (2) that have been isolated from a wide range of environments such as soil, rhizosphere, water, air, food, milk, human blood, and feces (3–5). During investigations on bacterial biodiversity in the soil sample of a pine garden, a new strain of the genus *Paenibacillus*, designated *Paenibacillus* sp. M.A.Huq-81 was isolated, and its draft genome assembly is presented in this study.

Strain M.A.Huq-81 was isolated from the soil sample of a pine garden, located in Anseong, South Korea (37° 03' 16" N, 127° 30' 59" E). One gram of soil sample (collection date: February 10, 2021) was suspended in 9 mL of sterile 0.85% (wt/vol) NaCl solution. The suspension was serially diluted up to a 10⁻⁶ dilution, and 100 µL of individual dilution was spread onto Tryptic Soy Agar (TSA) plates (6). Then, the plates were placed at 28 °C in an incubator for 3 days. Single colonies were purified by repeated streaking on fresh TSA plates (7). The strain M.A.Huq-81 has been deposited into the China General Microbiological Culture Collection Center with the deposition number CGMCC 1.60092. The genomic DNA was extracted from the bacterial culture that was grown in TSB for 24 h using the Solg Genomic DNA Prep kit (Solgent, Republic of Korea), according to the manufacturer's protocol. The DNA fragments were size selected using AMPure XP beads to isolate those within the desired range, typically 200–300 base pairs (bp) for standard Illumina sequencing libraries. The library for sequencing was prepared using the Nextera XT DNA Library Prep Kit (Illumina, San Diego, USA). After library preparation, sequencing was carried out using the Illumina HiSeq 3000 platform with paired-end 2 × 150 bp reads following the standard protocol provided by the manufacturer. Illumina's bcl2fastq software version 2.20.0 was utilized with default parameters to demultiplex the raw reads and convert them into FASTQ format. For further downstream analysis, raw paired-end reads from the HiSeq platform were used. Reads were processed with Trimmomatic version 0.38 (8). The genome sequence was assembled by SOAPdenovo v2.04 assembler (9). The genome was annotated with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 6.3 (10–12). The strain was identified using a whole genome-based taxonomic analysis in the Type (Strain) Genome Server (<https://tygs.dsmz.de>). The genome-to-genome distance (GGDH) bioinformatics tool (13) was

Editor Simon Roux, DOE Joint Genome Institute, Berkeley, California, USA

Address correspondence to Md. Shahedur Rahman, ms.rahman@just.edu.bd, or Jong-Whi Park, jpark@gachon.ac.kr.

The authors declare no conflict of interest.

Received 16 October 2025

Accepted 23 November 2025

Published 16 December 2025

Copyright © 2025 Huq 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 Genomic features of *Paenibacillus* sp. M.A.Huq-81

Feature	Result
Description of source	
Location	Anseong, South Korea
Time	2021
Type	Soil sample of pine garden
Sequencing summary	
Coverage	79×
Total bases	5,210,425
GC content	48.0%
Assembly report	
Contigs	22
Contig L_{50}	3
Contig N_{50}	566.4 kb
Genome length	5.2 Mb
Annotation report	
Genes (total)	4,913
CDSs (total)	4,822
CDSs (with protein)	4,797
ncRNAs	4
tRNA	80
Quality analysis	
Completeness	99.17%
Contamination	0.3%

used to measure *in silico* DNA-DNA hybridization (isDDH) values. The best matches to the strain M.A.Huq-81 genome are the type strain *Paenibacillus pinistramenti* ASL46 (accession [GCA_005869875.1](https://www.ncbi.nlm.nih.gov/nuclot/GCA_005869875.1)) with *in silico* DNA-DNA hybridization values of 22.4%.

The number of reads generated by genome sequencing was 3,163,908. The genome of *Paenibacillus* sp. M.A.Huq-81 is 5,210,425 bp long with a GC content of 48.0%. The assembly contains 22 contigs with a coverage of 79×. A total of 4,913 genes were predicted, of which 4,797 are protein-coding genes. The details of genomic features are presented in Table 1.

ACKNOWLEDGMENTS

No external funding was received for this study.

We give special thanks to the World Data Centre for Microorganisms (WDCM), The Institute of Microbiology of the Chinese Academy of Sciences, Beijing, China, for the draft genome sequencing of strain M.A.Huq-81.

Md. Amdadul Huq, Conceptualization, Formal analysis, Writing – original draft, Writing – review and editing | Jong-Whi Park, Writing – review and editing, | Md. Shahedur Rahman, Writing – review and editing.

AUTHOR AFFILIATIONS

¹Department of Life Sciences, College of BioNano Technology, Gachon University, Seongnam, Republic of Korea

²Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Jashore, Bangladesh

AUTHOR ORCID*s*

Md. Amdadul Huq  <http://orcid.org/0000-0002-4831-6368>

Jong-Whi Park  <http://orcid.org/0000-0002-9890-9241>

Md. Shahedur Rahman  <http://orcid.org/0009-0005-2028-4885>

AUTHOR CONTRIBUTIONS

Md. Amdadul Huq, Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review and editing | Jong-Whi Park, Writing – review and editing, Investigation | Md. Shahedur Rahman, Supervision, Writing – review and editing

DATA AVAILABILITY

The draft genome sequence of *Paenibacillus* sp. M.A.Huq-81 was deposited in GenBank under accession number [JBRFIM000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JBRFIM000000000). The raw reads are available in SRA under accession number [SRR33604962](https://www.ncbi.nlm.nih.gov/sra/SRR33604962). The version described in this paper is the first version, [JBRFIM000000000.1](https://www.ncbi.nlm.nih.gov/nuclseq/JBRFIM000000000.1). The data have been deposited in GenBank under BioSample number [SAMN51699928](https://www.ncbi.nlm.nih.gov/biosample/SAMN51699928) and BioProject number [PRJNA1331960](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1331960). The genome assembly is available under the accession [GCA_052931865.1](https://www.ncbi.nlm.nih.gov/nuclseq/GCA_052931865.1).

REFERENCES

- Ash C, Priest FG, Collins MD. 1993. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. proposal for the creation of a new genus *Paenibacillus*. *Antonie Van Leeuwenhoek* 64:253–260. <https://doi.org/10.1007/BF00873085>
- LPSN. 2025. Genus *Paenibacillus*. Available from: <https://lpsn.dsmz.de/genus/paenibacillus>
- Huq MA. 2020. *Paenibacillus anseongense* sp. nov. a silver nanoparticle producing bacterium isolated from rhizospheric soil. *Curr Microbiol* 77:2023–2030. <https://doi.org/10.1007/s00284-020-02086-0>
- Akter S, Huq MA. 2018. Biological synthesis of ginsenoside Rd using *Paenibacillus horti* sp:1566–1573. <https://doi.org/10.1007/s00284-018-1561-6>
- Huq MA, Kim YJ, Hoang VA, Siddiqi MZ, Yang DC. 2015. *Paenibacillus ginsengiterrae* sp. nov., a ginsenoside-hydrolyzing bacteria isolated from soil of ginseng field. *Arch Microbiol* 197:389–396. <https://doi.org/10.1007/s00203-014-1073-0>
- Huq MA, Kim YJ, Rahman MM, Alam MM, Srinivasan S, Kang KK, Akter S. 2025. *Vogesella oryzagri* sp. nov., isolated from the rhizosphere of rice and *in silico* genome mining for the prediction of biosynthetic gene clusters. *Int J Syst Evol Microbiol* 75:006687. <https://doi.org/10.1099/ijsem.0.006687>
- Huq MA, Rahman MS, Rahman MM. 2024. *Aquicola agrisoli* sp. nov., isolated from rhizospheric soil of eggplant and *in silico* genome mining for the prediction of biosynthetic gene clusters. *Int J Syst Evol Microbiol* 74:006355. <https://doi.org/10.1099/ijsem.0.006355>
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. <https://doi.org/10.1186/2047-217X-1-18>
- Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 49:D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
- Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>
- 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>
- Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. 2022. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 50:D801–D807. <https://doi.org/10.1093/nar/gkab902>