

Genome sequence of *Flavobacterium* sp. MAHUQ-57 isolated from the soil sample of a rice field located in Dighalgram, Magura, Bangladesh

Md. Amdadul Huq,¹ M. Mizanur Rahman,² Md. Shahedur Rahman³

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT This study reports the draft genome sequence of *Flavobacterium* sp. MAHUQ-57, a strain isolated from the soil sample of a rice field located in Dighalgram, Magura, Bangladesh, in 2019. The genome consists of a 3,472,804 base pair chromosome assembled into 23 contigs, encoding 3,202 predicted protein-coding genes.

KEYWORDS genome sequence, *Flavobacterium* sp. MAHUQ-57, rice field

Flavobacterium is a genus of gram-negative, nonmotile and motile, rod-shaped bacteria that consists of 330 validly published species (<https://lpsn.dsmz.de/genus/flavobacterium>). Members of the genus *Flavobacterium* were isolated from different habitats, such as plants, animals, algae, seawater, freshwater, soil, rhizosphere, compost, sediments, lagoons, Antarctic lakes, glaciers, activated sludge, water reservoirs, and wastewater treatment plant (1–4). During investigations on the biodiversity of bacteria in the rice field, a strain of genus *Flavobacterium*, designated as *Flavobacterium* sp. MAHUQ-57 was isolated, and its draft genome assembly is presented in this study.

Strain MAHUQ-57 was isolated from the soil sample of a rice field, located in Dighalgram, Magura, Bangladesh (GPS map: 23.410203, 89.332807). Soil samples were collected from the rhizosphere surrounding the plant roots. One gram of soil sample 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 R2A agar plates. Then, the plates were placed at a 30°C incubator for 3 days. Single colonies were purified by repeated streaking on fresh R2A agar plates. The genomic DNA was extracted from the bacterial culture that was grown in R2A broth for 24 h using Solg Genomic DNA Prep kit (Solgent, Republic of Korea), according to the manufacturer's protocol. DNA was fragmented using a tagmentation enzyme (Nextera XT kit, Illumina, San Diego, USA). 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). 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. Reads were quality-controlled using FastQC (5), and adapter sequences and low-quality bases were trimmed with Trimmomatic version 0.38 (6). The genome sequence was assembled by SOAPdenovo v2.04 assembler (7). The genome was annotated with the NCBI Prokaryotic Genome Annotation Pipeline version 6.3 (8–10). The completeness and contamination were assessed by CheckM tool, version 1.2.4 (11). 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

Editor John J. Dennehy, Queens College
Department of Biology, Queens, New York, USA

Address correspondence to Md. Shahedur Rahman,
ms.rahman@just.edu.bd.

The authors declare no conflict of interest.

Received 24 August 2025

Accepted 5 November 2025

Published 13 January 2026

Copyright © 2026 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 *Flavobacterium* sp. MAHUQ-57

Feature	Result
Description of source	
Location	Dighalgram, Magura, Bangladesh
Time	2019
Type	Soil of rice field
Sequencing summary	
Coverage	139×
Total bases	3,472,804
GC content	35.5%
Assembly report	
Contigs	23
Contig <i>L</i> ₅₀	3
Contig <i>N</i> ₅₀	653.8 kb
Genome length	3.5 Mb
Annotation report	
Genes (total)	3,262
CDSs (total)	3,209
CDSs (with protein)	3,202
ncRNAs	3
tRNA	44
Quality analysis	
Completeness	99.22%
Contamination	0.17%

bioinformatics tool (12) was used to measure *in silico* DNA-DNA hybridization values. The best matches to the strain MAHUQ-57 genome are the type strain *Flavobacterium amniphilum* KYPY10 (accession [NZ_JAMLJL000000000](https://www.ncbi.nlm.nih.gov/nuclot/NZ_JAMLJL000000000)) with *in silico* DNA-DNA hybridization values of 22.4%.

The number of reads generated by genome sequencing was 3,072,309. The genome of *Flavobacterium* sp. MAHUQ-57 is 3,472,804 bp long with a GC content of 35.5%. The assembly contains 23 contigs with a coverage of 139×. A total of 3,262 genes were predicted, of which 3,202 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 CGM 10K project for analyzing the draft genome sequence of strain MAHUQ-57 (GCM60027622).

M.A.H.: Conceptualization, Formal analysis, Writing—original draft, Writing—review and editing. M.M.R.: Writing—review and editing. M.S.R.: Writing—review and editing.

AUTHOR AFFILIATIONS

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

²Department of Biotechnology and Genetic Engineering, Faculty of Biological Science, Islamic University, Kushtia, Bangladesh

³Department of Genetic Engineering and Biotechnology, Jashore University of Science and Technology, Jashore, Bangladesh

AUTHOR ORCIDs

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

M. Mizanur Rahman  <http://orcid.org/0000-0003-4225-9451>

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

DATA AVAILABILITY

The draft genome sequence of *Flavobacterium* sp. MAHUQ-57 was deposited in GenBank under the accession number [JBPZTE000000000](#). The raw reads are available in SRA under the accession number [SRR33604967](#). The version described in this paper is the first version, JBPZTE000000000.1. The data have been deposited in GenBank under the BioSample number [SAMN50309237](#) and the BioProject number [PRJNA1299606](#). The genome assembly is available under the accession number [GCA_051908335.1](#).

REFERENCES

1. Akter S, Lee SY, Moon B, Huq MA. 2021. *Flavobacterium agri* sp. nov., a novel bacterial species isolated from rhizospheric soil of *Coriandrum sativum*. Arch Microbiol 203:701–706. <https://doi.org/10.1007/s00203-020-02068-7>
2. Liu Q, Siddiqi MZ, Liu Q, Huq MA, Lee SY, Choi KD, Im WT. 2018. *Flavobacterium hankyongi* sp. nov., isolated from activated sludge. Int J Syst Evol Microbiol 68:1732–1736. <https://doi.org/10.1099/ijsem.0.002739>
3. Huq MA, Akter S, Lee SY. 2018. *Flavobacterium chungangensis* sp. nov., a bacterium isolated from soil of Chinese cabbage garden. Curr Microbiol 75:842–848. <https://doi.org/10.1007/s00284-018-1456-6>
4. Huq MA, Akter S, Rahman MM, Parvez MAK, Alam MM, Rahman MS. 2025. Draft genome sequence of *Flavobacterium* sp. MAHUQ-50 isolated from the rhizosphere of banana in Magura, Bangladesh. Microbiol Resour Announc 14:e0062325. <https://doi.org/10.1128/mra.00623-25>
5. FastQC: a quality control tool for high throughput sequence data [Online]. 2010. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
6. 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>
7. 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>
8. 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>
9. 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>
10. 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>
11. 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>
12. 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>