




Draft Genome Sequence of *Bacillus marisflavi* CK-NBRI-03, Isolated from Agricultural Soil

Mayank Gupta,^a  Puneet Singh Chauhan,^b Sudhir K. Sopory,^a Sneha L. Singla-Pareek,^a Nidhi Adlakha,^c Ashwani Pareek,^d  Charanpreet Kaur^d

^aInternational Centre for Genetic Engineering and Biotechnology, New Delhi, India

^bMicrobial Technologies Division, CSIR-National Botanical Research Institute, Lucknow, Uttar Pradesh, India

^cRegional Centre for Biotechnology, NCR Biotech Science Cluster, Faridabad, Haryana, India

^dSchool of Life Sciences, Jawaharlal Nehru University, New Delhi, India

ABSTRACT Here, we report the 4.34-Mb draft genome assembly of *Bacillus marisflavi* CK-NBRI-03 (or P3), a Gram-positive bacterium, with an average G+C content of 48.66%. P3 was isolated from agricultural soil from the Badaun (midwestern plain zone) region of Uttar Pradesh, India.

Bacillus marisflavi CK-NBRI-03 (P3) was isolated from a wheat field located in the Badaun region of Uttar Pradesh, India, while the microbial diversity of that area was explored post-wheat harvest. Preliminary 16S rRNA gene sequencing studies using the 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-TACGGYTACCTGTTACGACTT-3') primers (1) revealed 99.77% similarity (with 90% query coverage) of P3 to *Bacillus marisflavi* TF-11 (JCM 11544), a carotenoid-producing bacterium isolated from seawater (2). Although bacteria belonging to the *Bacillus* genus are well-known plant growth promoters, information regarding the involvement of any *B. marisflavi* strain in plant growth promotion is lacking. Therefore, sequencing of the P3 genome was undertaken to investigate its genomic features and to assess the potential role of this isolated *B. marisflavi* strain in influencing plant growth.

P3 was isolated from agricultural soil per a protocol described previously (3). For whole-genome sequencing studies, the strain was preserved as glycerol stock after initial isolation and purification and then restreaked onto a nutrient agar plate, and a single colony was inoculated in the nutrient broth. The culture was then grown at 28°C for 24 h, and harvested cells were subsequently used for genomic DNA isolation, carried out using the GenElute bacterial genomic DNA kit (Sigma-Aldrich). This was followed by library preparation using the NEBNext Ultra DNA library prep kit for Illumina (New England Biolabs, Ipswich, MA) as per the manufacturer's instructions.

The Illumina HiSeq 2500 platform with 100-bp paired-end reads was used for the sequencing of the P3 genome, which generated a total of 949.46 Mb of raw reads. From these, Illumina adapter sequences were removed using Cutadapt version 1.14 (4). Low-quality (Q < 30) reads were filtered out using Sickle version 1.33 (5), and duplicate reads were removed using FastUniq 1.1 (6). After preprocessing, we obtained 857.40 Mb of clean paired-end reads at ~200× genome coverage and with an average DNA G+C content of 48.66%. The reads were separately assembled using Velvet version 1.2.10 (7) and MaSuRCA version 2.2.1 (8) tools. The resulting assemblies were subsequently merged using GAA version 1.0 (9). Thereafter, PAGIT version 1 (10) was used for the scaffolding of the merged assembly. Sixteen scaffolds containing a total of 4,344,737 bp with an N_{50} value of 696,266 bp were obtained. The average scaffold length was 271,546 bp, and the longest and shortest scaffolds were 1,820,907 bp and 1,001 bp, respectively.

Citation Gupta M, Chauhan PS, Sopory SK, Singla-Pareek SL, Adlakha N, Pareek A, Kaur C. 2020. Draft genome sequence of *Bacillus marisflavi* CK-NBRI-03, isolated from agricultural soil. Microbiol Resour Announc 9:e00044-20. <https://doi.org/10.1128/MRA.00044-20>.

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Copyright © 2020 Gupta 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/).

Address correspondence to Charanpreet Kaur, charanpreet06@gmail.com.

Received 16 January 2020

Accepted 21 January 2020

Published 13 February 2020

The final draft genome was then annotated using the standalone Prokaryotic Genome Annotation Pipeline (PGAP) version 2019-08-01.build3919 (11). A total of 4,422 genes were predicted, including 4,280 coding sequences (CDS) and 142 RNA genes. Among the RNA genes, 25 were rRNA genes (9 5S rRNA, 8 16S rRNA, and 8 23S rRNA), and 112 were tRNA genes. Protein sequences were annotated using the BlastKOALA (12) tool (accessed January 2019) for Kyoto Encyclopedia of Genes and Genomes (KEGG) enzyme codes. The proteins were also clustered on the basis of homology using the annotation resource Clusters of Orthologous Groups (COG) database (accessed January 2019) (13). In total, 2,219 and 2,910 genes were assigned to KEGG orthology (KO) and COG categories, respectively. About 1,275 protein-coding genes were connected to KEGG pathways, and 1,172 genes encoded enzymes mapping to Enzyme Classification numbers. Default parameters were used to run all the software/tools unless otherwise specified.

Future investigations may delineate the role of *B. marisflavi* strain P3 as a potential biofertilizer, similar to other members of the *Bacillus* genus.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [VSJG00000000](https://www.ncbi.nlm.nih.gov/nuclseq/VSJG00000000). The version described in this paper is version VSJG00000000.1. The BioProject accession number is [PRJNA478293](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA478293).

ACKNOWLEDGMENT

C.K. acknowledges DST-INSPIRE grant IFA-14/LSPA-24, received from the Department of Science and Technology (DST), Government of India.

REFERENCES

- Lane DJ. 1991. 16S/23S rRNA sequencing, p 115–176. In Stackebrandt E, Goodfellow M (ed), *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, Inc., New York, NY.
- Wang JP, Liu B, Liu GH, Chen DJ, Chen QQ, Zhu YJ, Chen Z, Che JM. 2015. Draft genome sequence of *Bacillus marisflavi* TF-11T (JCM 11544), a carotenoid-producing bacterium isolated from seawater from a tidal flat in the Yellow Sea. *Genome Announc* 3:e01451-15. <https://doi.org/10.1128/genomeA.01451-15>.
- Gupta M, Chauhan PS, Sopory SK, Singla-Pareek SL, Pareek A, Adlakha N, Kaur C. 2019. Draft genome sequence of a potential plant growth-promoting rhizobacterium, *Pseudomonas* sp. strain CK-NBRI-02. *Microbiol Resour Announc* 8:e01113-19. <https://doi.org/10.1128/MRA.01113-19>.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
- Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). <https://github.com/najoshi/sickle>.
- Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S. 2012. FastUniq: a fast *de novo* duplicates removal tool for paired short reads. *PLoS One* 7:e52249. <https://doi.org/10.1371/journal.pone.0052249>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>.
- Yao G, Ye L, Gao H, Minx P, Warren WC, Weinstock GM. 2012. Graph concordance of next-generation sequence assemblies. *Bioinformatics* 28:13–16. <https://doi.org/10.1093/bioinformatics/btr588>.
- Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc* 7:1260–1284. <https://doi.org/10.1038/nprot.2012.068>.
- Haft DH, DiCuccio M, Badretdin A, Brover V, Chetverin 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>.
- Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 428:726–731. <https://doi.org/10.1016/j.jmb.2015.11.006>.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36. <https://doi.org/10.1093/nar/28.1.33>.