



# Draft Genome Sequence of *Bacillus* sp. Strain IGA-FME-2, Isolated from the Bulk Soil of Soybean (*Glycine max* L.) in Northeast China

Zhenhua Yu,<sup>a</sup> Sergio de los Santos-Villalobos,<sup>b</sup> Yansheng Li,<sup>a</sup> Jian Jin,<sup>a</sup> Fannie Isela Parra Cota,<sup>c</sup> Valeria Valenzuela Ruiz,<sup>b</sup> Guanghua Wang<sup>a</sup>

<sup>a</sup>Key Laboratory of Mollisols Agroecology, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China

<sup>b</sup>Instituto Tecnológico de Sonora, Ciudad Obregón, Sonora, Mexico

<sup>c</sup>Campo Experimental Norman E. Borlaug, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Ciudad Obregón, Sonora, Mexico

**ABSTRACT** Here, we present the draft genome of *Bacillus* sp. strain IGA-FME-2. This strain was isolated from the bulk soil of soybean (*Glycine max* L.). Its genome consists of 3,810 protein-coding genes, 44 tRNAs, two 16S rRNAs, and a single copy of 23S rRNA, with a GC content of 46.4%.

Soybean (*Glycine max* L.) is cultivated worldwide due to its nutritional value and oil-yielding characteristics (1). Successful soybean production is hampered by abiotic and biotic stresses such as drought, weeds, insect pests, and diseases, which reduce yields by up to 40% and increase the cost of cultivation (2, 3). Thus, sustainable agricultural practices, such as the application of plant growth-promoting bacteria (4–8), are needed to combat these stresses.

Strain IGA-FME-2 was isolated from the bulk soil of soybean in an agricultural field in Harbin, Heilongjiang, China (45°41'N, 126°38'E). One gram of collected soil was suspended in 9 ml of sterile (treated at 121°C and 15 lb/in<sup>2</sup> for 15 min) distilled water and homogenized for 15 min at 5 × *g* (using a rotary shaker). The soil suspension was serially diluted from 10<sup>-1</sup> to 10<sup>-6</sup>, 0.1 ml of each dilution was pipetted out and spread onto plates with solid LB medium, and the plates were incubated at 28°C for 48 h. Bacterial colonies were purified by sequential streaking onto solid LB medium, and macroscopic/microscopic characterization was carried out (in triplicate) to ensure axenicity. All isolates obtained were cryopreserved at -80°C using liquid LB medium and 30% glycerol. Then, high-quality genomic DNA was extracted and purified from strain IGA-FME-2, after growth in liquid LB medium at 28°C for 48 h, by using the E.Z.N.A. bacterial DNA kit (Omega, USA) according to the manufacturer's instructions. The purified DNA was quantified with a TBS-380 fluorometer (Turner BioSystems, Inc., Sunnyvale, CA). High-quality DNA (optical density at 260 nm [OD<sub>260</sub>]/OD<sub>280</sub>, 1.8 to 2.0; total amount of DNA, ≥1 μg; concentration, ≥20 ng/μl) was used for sequencing with the HiSeq 2000 (2 × 100 bp) platform (Illumina, USA). Next-generation sequencing library preparation was carried out by using the NEXTflex rapid DNA-Seq kit version 2.0 for Illumina platforms, according to the manufacturer's instructions. The quality of the raw reads obtained was analyzed by FastQC version 0.11.5 (9), and all parameters showed very good quality except for sequence duplication levels and per-base sequence content, which showed intermediate quality. Trimmomatic version 0.32 was used to remove adapter sequences by using the universal adapter sequences, low-quality bases by using a sliding window approach of 4:24, and bases at the start of the read, according to the sequencing quality, by using HEADCROP. With these parameters, only 1.95% of reads were dropped, and after a FastQC reanalysis, all parameters were significantly improved to very good quality. Subsequently, a *de novo* assembly was generated by

**Citation** Yu Z, de los Santos-Villalobos S, Li Y, Jin J, Parra Cota FI, Valenzuela Ruiz V, Wang G. 2021. Draft genome sequence of *Bacillus* sp. strain IGA-FME-2, isolated from the bulk soil of soybean (*Glycine max* L.) in Northeast China. Microbiol Resour Announc 10:e00004-21. <https://doi.org/10.1128/MRA.00004-21>.

**Editor** Julia A. Maresca, University of Delaware

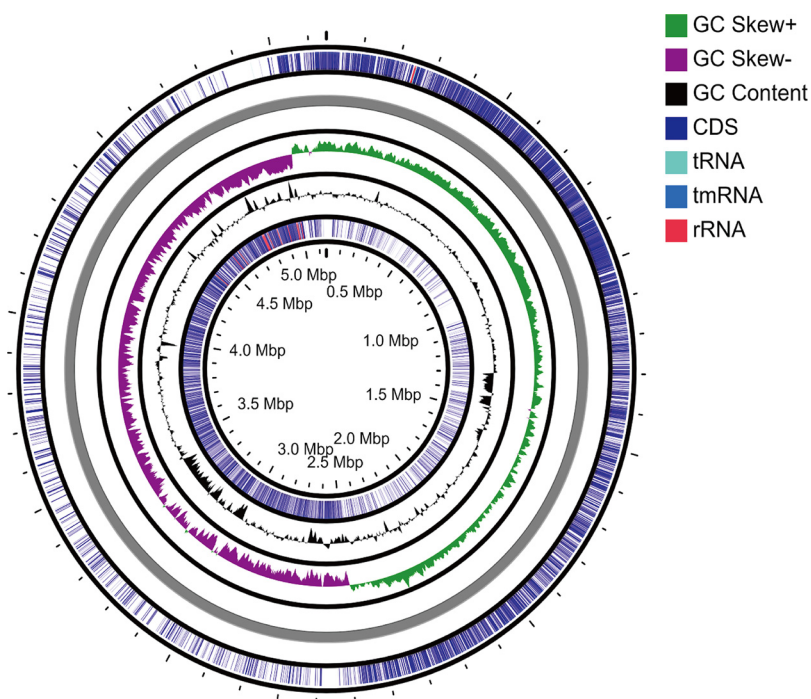
**Copyright** © 2021 Yu et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Guanghua Wang, wanggh@iga.ac.cn.

**Received** 5 January 2021

**Accepted** 12 March 2021

**Published** 22 April 2021



**FIG 1** Circular chromosome map of *Bacillus* sp. strain IGA-FME-2, showing the distribution of coding DNA sequences (CDS), tRNAs, rRNAs, and GC content skew (50% of the total base pair window).

SPAdes version 3.14.1 (10), using the parameter `-careful` for error correction in reads and `-cov-cutoff auto`. The draft genome of strain IGA-FME-2 consisted of 40 contigs (>200 bp) (minimum, 209 bp; maximum, 984,802 bp;  $N_{50}$ , 436,298 bp;  $L_{50}$ , 3). The final assembly contained 3,955,933 bp, with a GC content of 46.4%. The assembled contigs were ordered by Mauve Contig Mover version 2.4.0 (11), using the reference genome of *Bacillus velezensis* FZB42 (GenBank accession number [GCA\\_000015785.1](https://doi.org/10.1093/nar/gkz111)). The reference genome was selected based on the greatest 16S rRNA identity (100%). The circular chromosome map was then generated using the CGView Server (12) (Fig. 1).

Genome annotation was performed by Prokka v 1.11.0 (13). The genome is predicted to contain 3,810 protein-coding genes, 44 tRNAs, 2 16S rRNAs, and a single copy of 23S rRNA.

**Data availability.** This draft genome sequence has been deposited in DDBJ/ENA/GenBank under accession number [JADDHM000000000](https://doi.org/10.1093/nar/gkz111). The version described in this paper is the first version, [JADDHM010000000](https://doi.org/10.1093/nar/gkz111), under BioProject number [PRJNA668193](https://doi.org/10.1093/nar/gkz111) and BioSample number [SAMN16400875](https://doi.org/10.1093/nar/gkz111). Raw data have been deposited in the NCBI SRA under accession number [PRJNA675398](https://doi.org/10.1093/nar/gkz111) (sample IGA-FME-2 [BioSample number [SAMN16706821](https://doi.org/10.1093/nar/gkz111) and SRA accession number [SRX9460425](https://doi.org/10.1093/nar/gkz111)]).

## ACKNOWLEDGMENTS

This research was funded by the Youth Innovation Promotion Association of the Chinese Academy of Sciences (grant 2019233) and the Scientific Research Foundation for the Returned Overseas Chinese Scholars of Heilongjiang Province.

## REFERENCES

- Murithi HM, Beed F, Tukamuhabwa P, Thomma BPHJ, Joosten MHAJ. 2016. Soybean production in eastern and southern Africa and threat of yield loss due to soybean rust caused by *Phakopsora pachyrhizi*. *Plant Pathol* 65:176–188. <https://doi.org/10.1111/ppa.12457>.
- Chung G, Singh RJ. 2008. Broadening the genetic base of soybean: a multidisciplinary approach. *Crit Rev Plant Sci* 27:295–341. <https://doi.org/10.1080/07352680802333904>.
- Tewari S, Arora NK, Miransari M. 2016. Plant growth promoting rhizobacteria to alleviate soybean growth under abiotic and biotic stresses, p 131–155. In Miransari M (ed), *Abiotic and biotic stresses in soybean production: soybean production*, vol 1. Academic Press, San Diego, CA. <https://doi.org/10.1016/B978-0-12-801536-0.00006-2>.
- de los Santos Villalobos S, Parra Cota FI, Herrera Sepúlveda A, Valenzuela Aragón B, Estrada Mora JC. 2018. *Colmena: colección de*

- microorganismos edáficos y endófitos nativos, para contribuir a la seguridad alimentaria nacional. *Remexca* 9:191–202. <https://doi.org/10.29312/remexca.v9i1.858>.
5. Villarreal-Delgado MF, Villa-Rodríguez ED, Cira-Chávez LA, Estrada-Alvarado MI, Parra-Cota FI, de los Santos-Villalobos S. 2018. The genus *Bacillus* as a biological control agent and its implications in the agricultural biosecurity. *Rev Mex Fitopatol* 36:95–130. <https://doi.org/10.18781/R.MEX.FIT.1706-5>.
  6. de los Santos-Villalobos S, Robles RI, Parra Cota FI, Larsen J, Lozano P, Tiedje MJ. 2019. *Bacillus cabrialesii* sp. nov., an endophytic plant growth promoting bacterium isolated from wheat (*Triticum turgidum* subsp. durum) in the Yaqui Valley, Mexico. *Int J Syst Evol Microbiol* 69:3939–3945. <https://doi.org/10.1099/ijsem.0.003711>.
  7. Yasmin H, Naeem S, Bakhtawar M, Jabeen Z, Nosheen A, Naz R, Keyani R, Mumtaz S, Hassan MN. 2020. Halotolerant rhizobacteria *Pseudomonas pseudoalcaligenes* and *Bacillus subtilis* mediate systemic tolerance in hydroponically grown soybean (*Glycine max* L.) against salinity stress. *PLoS One* 15:e0231348. <https://doi.org/10.1371/journal.pone.0231348>.
  8. Valenzuela-Aragon B, Parra-Cota FI, Santoyo G, Arellano-Wattenbarger GL, de los Santos-Villalobos S. 2019. Plant-assisted selection: a promising alternative for in vivo identification of wheat (*Triticum turgidum* L. subsp. *Durum*) growth promoting bacteria. *Plant Soil* 435:367–384. <https://doi.org/10.1007/s11104-018-03901-1>.
  9. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
  10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
  11. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignments of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>.
  12. Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res* 36:W181–W184. <https://doi.org/10.1093/nar/gkn179>.
  13. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.