



Genome Sequence of *Bacillus subtilis* subsp. *subtilis* Strain IITK SM1, Isolated from Kitchen Waste Compost

Murugan Prem Anand,^a V. Chinnasamy Hariharan,^c Hari Shankar,^b Abhas Singh,^{b,c}  Matheshwaran Saravanan^{a,c}

^aDepartment of Biological Sciences and Bioengineering, Indian Institute of Technology, Kanpur, India

^bDepartment of Civil Engineering, Indian Institute of Technology, Kanpur, India

^cCentre for Environmental Sciences and Engineering, Indian Institute of Technology, Kanpur, India

ABSTRACT We report here the complete genome sequence of *Bacillus subtilis* subsp. *subtilis* strain IITK SM1, isolated from kitchen waste compost. We have sequenced the whole genome of this strain to identify and characterize the enzymes that participate in efficient composting activity.

Bacillus subtilis is a rod-shaped Gram-positive bacterium which has been used as a model organism for several years. *B. subtilis*, present in different environmental niches, is known to evolve to circumvent the deleterious effects of chemicals and other contaminants (1). *B. subtilis* strains are known to produce extracellular enzymes that are involved in several biological processes. These enzymes help the bacteria adapt to environmental changes and also protect the plant in their environment; hence, it is exploited as a biocontrol agent (2). Different strains of *B. subtilis* have potential applications in the management of waste treatment, prevention of plant disease, composting, and bioremediation (3). We have isolated and characterized a *B. subtilis* subsp. *subtilis* strain, IITK SM1, from kitchen compost in Kanpur, India. The strain was found to have efficient composting activity of kitchen waste, which will be reported elsewhere.

We isolated a single colony of the strain by serial dilutions, followed by the streak plate method. The strain was grown overnight in 10 ml of Luria broth at 37°C. Genomic DNA was purified via a phenol-chloroform extraction method. The quantity and quality of isolated DNA were determined using a NanoDrop spectrophotometer. The paired-end (PE) sequencing library of the sample was prepared using a TruSeq Nano DNA library prep kit. A PCR-enriched library was analyzed on a 4200 TapeStation system (Agilent Technologies) using a high-sensitivity D1000 ScreenTape assay. The mean fragment size of the PCR-enriched library was found to be 498 bp with concentration of 816 pg/μl with region molarity of 2,590 pM and 90.61% in total. The PE Illumina library was loaded onto a NextSeq 500 system for cluster generation and sequencing (Eurofins). The library was sequenced using Illumina chemistry with 150-bp paired-end reads and 2 × 150-bp chemistry. The sequenced raw data were processed to obtain high-quality clean reads using Trimmomatic version 0.35 to remove adaptor sequences, ambiguous reads, and low-quality sequences (4). *De novo* assembly and scaffolding were done using Velvet (version 1.2.10) using an optimized k-mer of 59 (5). The scaffolds were merged using CONTIGuator (6). The reference genome used was that of *B. subtilis* subsp. *subtilis* strain NCD-2 (4.19 Mb in length). The final genome size was found to be 4,060,726 bp with a GC content of 43.58%. The genes were predicted in the assembled scaffolds using Prodigal (7). Gene ontology and functional annotations were performed using the Blast2GO platform (8). The total number of genes was 4,085, of which 1,390 genes were associated with various biological processes, 1,066 genes were associated with the formation of cellular components, and 1,438 genes were involved

Citation Prem Anand M, Hariharan VC, Shankar H, Singh A, Saravanan M. 2019. Genome sequence of *Bacillus subtilis* subsp. *subtilis* strain IITK SM1, isolated from kitchen waste compost. Microbiol Resour Announc 8:e01330-18. <https://doi.org/10.1128/MRA.01330-18>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2019 Prem Anand 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 Matheshwaran Saravanan, saran@iitk.ac.in.

Received 26 September 2018

Accepted 9 January 2019

Published 7 February 2019

in molecular functions. There were 20 genes without a BLAST hit, which were in the size range of 90 to 260 bp. The Circos plot was generated using Circoletto (9), with the closest reference genome being that of *Bacillus subtilis* subsp. *subtilis* strain NCD-2 (GenBank accession number CP023755), which showed the synteny and clear representation of intergenic relationships. Using KASS (10), a total of 1,772 genes were found to be involved in 24 different KEGG pathways. Several genes were also identified that are known to cause xenobiotic degradation.

The genome sequence of *B. subtilis* subsp. *subtilis* strain IITK SM1 is important for understanding the factors that help in efficient kitchen waste composting and secondary metabolite production.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number CP031675. The version described in this paper is version CP031675.1. The raw data have been registered in the NCBI SRA database under the accession number SRR8290379.

ACKNOWLEDGMENTS

We thank Eurofins Genomics India Pvt. Ltd. for sequencing and analysis of the genome and the Centre for Environmental Sciences and Engineering, IIT Kanpur for the lab facility.

REFERENCES

1. Macdonald CB, Stockbridge RB. 2017. A topologically diverse family of fluoride channels. *Curr Opin Struct Biol* 45:142–149. <https://doi.org/10.1016/j.sbi.2017.04.003>.
2. van Dijl JM, Hecker M. 2013. *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. *Microb Cell Fact* 12:3. <https://doi.org/10.1186/1475-2859-12-3>.
3. Gupta R, Beg QK, Lorenz P. 2002. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl Microbiol Biotechnol* 59: 15–32. <https://doi.org/10.1007/s00253-002-0975-y>.
4. 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>.
5. 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>.
6. Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med* 6:11. <https://doi.org/10.1186/1751-0473-6-11>.
7. Hyatt D, Chen G, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
8. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36:3420–3435. <https://doi.org/10.1093/nar/gkn176>.
9. Darzentas N. 2010. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics* 26:2620–2621. <https://doi.org/10.1093/bioinformatics/btq484>.
10. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 35:182–185. <https://doi.org/10.1093/nar/gkm321>.