





# Whole-Genome Sequence of *Bacillus megaterium* Strain SGAir0080, Isolated from an Indoor Air Sample

Namrata Kalsi,<sup>a</sup> Akira Uchida,<sup>a</sup>  Rikky W. Purbojati,<sup>a</sup> James N. I. Houghton,<sup>a</sup> Caroline Chénard,<sup>a</sup> Anthony Wong,<sup>a</sup> Megan E. Clare,<sup>a</sup> Kavita K. Kushwaha,<sup>a</sup> Alexander Putra,<sup>a</sup> Nicolas E. Gaultier,<sup>a</sup> Balakrishnan N. V. Premkrishnan,<sup>a</sup> Cassie E. Heinle,<sup>a</sup> Vineeth Kodengil Vettath,<sup>a</sup> Daniela I. Drautz-Moses,<sup>a</sup>  Ana Carolina M. Junqueira,<sup>b</sup> Stephan C. Schuster<sup>a</sup>

<sup>a</sup>Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore

<sup>b</sup>Departamento de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

**ABSTRACT** *Bacillus megaterium* strain SGAir0080 was isolated from a tropical air sample in Singapore. Its genome was assembled using single-molecule real-time (SMRT) sequencing and MiSeq reads. It has one chromosome of 5.06 Mbp and seven plasmids (average length, 62.8 kbp). It possesses 5,339 protein-coding genes, 130 tRNAs, and 35 rRNAs.

*Bacillus megaterium* (*Firmicutes*) is an aerobic, spore-forming, Gram-positive species (1) that was first described by De Bary in 1884 (2). In the field of biotechnology, *B. megaterium* is used in the production of intracellular recombinant proteins (3–6). Although *B. megaterium* is found predominantly in soil (7), it is ubiquitous in the environment and has been reported from air samples (8).

*B. megaterium* was sequenced here in an effort to investigate the diverse microbiological communities present in air. The *B. megaterium* strain SGAir0080 was isolated in Singapore (global positioning system [GPS] coordinates 1.347654 N, 103.685240 E) from an air sample collected using the Andersen single-stage impactor (SKC, Inc., USA) and impacted onto marine agar (Becton, Dickinson, USA). Colonies were isolated by growing cultures overnight on Trypticase soy agar (Becton, Dickinson) at 30°C, followed by cultivation overnight in lysogeny broth at 30°C prior to DNA extraction. DNA was purified using the Wizard genomic DNA purification kit (Promega, USA). A genomic DNA library was then prepared with the SMRTbell template prep kit 1.0 (Pacific Biosciences, USA). DNA was sheared using the g-Tube shearing method and size selected using BluePippin size selection (cutoff, 15 kb). This was followed by single-molecule real-time (SMRT) sequencing on a PacBio RS II sequencer (DNA sequencing kit 4.0 v2). Whole-genome shotgun libraries were constructed using the TruSeq Nano DNA library preparation kit (Illumina, USA), and short-read data were generated via a paired-end Illumina MiSeq run with a 300-bp read length. For the following analysis, all software was run with default settings unless otherwise stated.

Quality control of PacBio reads and MiSeq reads was performed using PreAssembler Filter v1 from the Hierarchical Genome Assembly Process v3 (HGAP3) (9) protocol, implemented in the PacBio SMRT Analysis 2.3.0 package and Cutadapt v1.8.1 (10), respectively. *De novo* assembly for 31,318 PacBio subreads ( $N_{50}$ , 16,423 bp) was performed using HGAP3 and polished with Quiver (9). The quality of the draft assembly was further improved with 697,928 MiSeq paired-end reads using Pilon v1.16 (11) (tracks –changes –vcf –fix all –mindepth 0.1 –mingap 10 –minmq 30 –minqual 20 –K 47). The complete assembly consists of 8 contigs (Table 1), with the chromosome having a G+C content of 38.2%. Contig lengths and G+C content were obtained with the Quality Assessment Tool for Genome Assemblies (QUAST) (12). Completeness and circularity of the chromosome and plasmids were evaluated using BUSCO (13) and Circlator v1.1.4 (14).

**Citation** Kalsi N, Uchida A, Purbojati RW, Houghton JN, Chénard C, Wong A, Clare ME, Kushwaha KK, Putra A, Gaultier NE, Premkrishnan BNV, Heinle CE, Vettath VK, Drautz-Moses DI, Junqueira ACM, Schuster SC. 2019. Whole-genome sequence of *Bacillus megaterium* strain SGAir0080, isolated from an indoor air sample. *Microbiol Resour Announc* 8:e01249-19. <https://doi.org/10.1128/MRA.01249-19>.

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

**Copyright** © 2019 Kalsi 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 Stephan C. Schuster, [SCSchuster@ntu.edu.sg](mailto:SCSchuster@ntu.edu.sg).

**Received** 3 October 2019

**Accepted** 7 November 2019

**Published** 12 December 2019

**TABLE 1** Lengths and coverages of strain SGAir0080 contigs

Contig name	Length (bp)	Coverage (×)	GenBank accession no.
SGAir0080 chromosome	5,057,175	50.1	CP028084
SGAir0080 unnamed_1	122,231	39.1	CP028085
SGAir0080 unnamed_2	140,900	21.3	CP028088
SGAir0080 unnamed_3	37,586	16.4	CP028086
SGAir0080 unnamed_4	60,004	19.8	CP028087
SGAir0080 unnamed_6	24,808	13.8	CP028089
SGAir0080 unnamed_7	19,487	13.6	CP028090
SGAir0080 unnamed_8	35,203	17.6	CP028091

NCBI's Prokaryotic Genome Annotation Pipeline (PGAP) v4.2 (15) was used for annotation, with a cutoff of 80% to determine the presence for a gene in the genome. This revealed 5,649 genes with 5,339 protein-coding genes, 35 rRNA genes (13 5S, 11 16S, and 11 23S), 130 tRNAs, 8 noncoding RNAs, and 137 pseudogenes.

Taxonomic identification was performed with Phyla-AMPHORA (16) using MarkerScanner.pl with an added “-DNA” flag and MarkerAlignTrim.pl with options “-WithReference” and “-OutputFormat phylip.” Phylotyping.pl was run with default parameters. SGAir0080 showed 99.3% identity with *B. megaterium* (minimum confidence, 1.0). Average nucleotide identity (ANI) analysis was performed with Microbial Species Identifier (MISI) (17) against a database of 6,387 bacterial RefSeq genomes with a text filter for “type,synonym type, proxytype” and subsequent “getorf -find 3” option. This gave 97.5% similarity and an alignment fraction value of 0.73 to *B. megaterium*.

**Data availability.** The genome sequence of *Bacillus megaterium* strain SGAir0080 and its plasmids have been deposited in DDBJ/EMBL/GenBank under accession numbers CP028084, CP028085, CP028086, CP028087, CP028088, CP028089, CP028090, and CP028091, respectively, and in the SRA database under accession numbers SRR8894398 and SRR8894399, respectively.

## ACKNOWLEDGMENT

The work was supported by a Singapore Ministry of Education Academic Research Fund tier 3 grant (MOE2013-T3-1-013).

## REFERENCES

- Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman W. 2009. The *Firmicutes*. In *Bergey's manual of systematic bacteriology*, vol 3. Springer, New York, NY.
- De Bary A. 1884. Vergleichende Morphologie und Biologie der Pilze Mycetozoen und Bacterien. Wilhelm Engelmann, Leipzig, Germany.
- Bäumchen C, Roth AH, Biedendieck R, Malten M, Follmann M, Sahm H, Bringer-Meyer S, Jahn D. 2007. D-Mannitol production by resting state whole cell biotransformation of D-fructose by heterologous mannitol and formate dehydrogenase gene expression in *Bacillus megaterium*. *Biotechnol J* 2:1408–1416. <https://doi.org/10.1002/biot.200700055>.
- Burger S, Tatge H, Hofmann F, Genth H, Just I, Gerhard R. 2003. Expression of recombinant *Clostridium difficile* toxin A using the *Bacillus megaterium* system. *Biochem Biophys Res Commun* 307:584–588. [https://doi.org/10.1016/s0006-291x\(03\)01234-8](https://doi.org/10.1016/s0006-291x(03)01234-8).
- Rygas T, Hillen W. 1991. Inducible high-level expression of heterologous genes in *Bacillus megaterium* using the regulatory elements of the xylose-utilization operon. *Appl Microbiol Biotechnol* 35:594–599. <https://doi.org/10.1007/bf00169622>.
- Rygas T, Scheler A, Allmansberger R, Hillen W. 1991. Molecular cloning, structure, promoters and regulatory elements for transcription of the *Bacillus megaterium* encoded regulon for xylose utilization. *Arch Microbiol* 155:535–542. <https://doi.org/10.1007/bf00245346>.
- Vary PS. 1994. Prime time for *Bacillus megaterium*. *Microbiology* 140: 1001–1013. <https://doi.org/10.1099/13500872-140-5-1001>.
- Zhiguo F, Chanjuan G, Zhiyun O, Peng L, Li S, Xiaoyong W. 2013. Characteristic and concentration distribution of culturable airborne bacteria in residential environments in Beijing, China. *Aerosol Air Qual Res* 14:943–953.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J* 17:1. <https://doi.org/10.14806/ej.17.1.200>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
- Felipe AS, Robert MW, Panagiotis L, Evgenia VK, Evgeny MZ. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 16:294. <https://doi.org/10.1186/s13059-015-0849-0>.
- Tatusova T, Dicuccio M, Badretdin A, Chetverin 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>.
- Wang Z, Wu M. 2013. A phylum-level bacterial phylogenetic marker database. *Mol Biol Evol* 30:1258–1262. <https://doi.org/10.1093/molbev/mst059>.
- Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpidis NC, Pati A. 2015. Microbial species delineation using whole genome sequences. *Nucleic Acids Res* 43:6761–6771. <https://doi.org/10.1093/nar/gkv657>.