

# Draft Genome Sequence of a Moderately Halophilic *Bacillus megaterium* Strain, MSP20.1, Isolated from a Saltern of the Little Rann of Kutch, India

Kamal Krishna Pal, Rinku Dey, Dharmesh Sherathia, Sejal Vanpariya, Ilaxi Patel, Trupti Dalsania, Kinjal Savsani, Bhoomika Sukhadiya, Mona Mandaliya, Manesh Thomas, Sucheta Ghorai, Rupal Rupapara, Priya Rawal, Abhi Shah, Sharmila Bhayani

Microbiology Section, Directorate of Groundnut Research, Junagadh, Gujarat, India

**The 4.37-Mbp draft genome of a moderately halophilic *Bacillus megaterium* strain, MSP20.1, isolated from a saltern of the Little Rann of Kutch, India, is reported here. To understand the mechanism(s) of moderate halophilism and to isolate the gene(s) involved in osmotolerance and adaptation, the genome of MSP20.1 was sequenced.**

Received 27 November 2013 Accepted 2 December 2013 Published 9 January 2014

**Citation** Pal KK, Dey R, Sherathia D, Vanpariya S, Patel I, Dalsania T, Savsani K, Sukhadiya B, Mandaliya M, Thomas M, Ghorai S, Rupapara R, Rawal P, Shah A, Bhayani S. 2014. Draft genome sequence of a moderately halophilic *Bacillus megaterium* strain, MSP20.1, isolated from a saltern of the Little Rann of Kutch, India. *Genome Announc*. 2(1):e01134-13. doi:10.1128/genomeA.01134-13.

**Copyright** © 2014 Pal et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Kamal Krishna Pal, [kkpal9426476749@gmail.com](mailto:kkpal9426476749@gmail.com).

The genomes of a number of species of *Bacillus* which can thrive in the salinity gradients (0 to 35% NaCl) of the hypersaline habitats of the Great and Little Rann of Kutch, India, were sequenced recently with a view to understanding the mechanism(s) of osmotolerance (1–5). *Bacillus megaterium* strain MSP20.1 (16S rRNA GenBank accession number JF802191), isolated from a saltern of the Little Rann of Kutch, India, grows optimally at a concentration of 7.5% NaCl (range, 0 to 20%) in the growth medium, at pH 7.0 and 37°C. The genome of MSP20.1 was sequenced to gain understanding of its moderate halophilism.

The genome of MSP20.1 was sequenced using the Illumina HiSeq 2000 platform at Macrogen, Inc., South Korea, through Sequencher Tech Pvt., Ltd., Ahmedabad, India. Originally, a total of 69,373,018 paired-end reads with lengths of 72 nucleotides of 7,006,674,818 bases were generated, which, after filtering and removal of PCR duplicates, resulted in 61,462,320 reads of 6,207,694,320 bases.

The genome was assembled, *de novo*, by use of SOAPdenovo assembler v 1.05 (6). The long assembly and the order of the contigs were generated by OSLay (7), with genome coverage of about 1,200×. The draft assembly of MSP20.1 (G+C content of 36.47%) resulted in an approximate genome size of 4,363,086 bases. The draft genome consists of 78 scaffolds of 3,658,318 bases, with the remaining bases included as unplaced scaffolds, with an average scaffold size of 46,901 bp (minimum, 1,057 bp; maximum, 357,224 bp) and  $N_{50}$  scaffold lengths of 65,811 bp. The assembly consists of 416 contigs (292 with lengths >500 bp), 360 being in the scaffolds with an average contig length of 10,488 bp (maximum, 95,529 bp; minimum, 55 bp) and  $N_{50}$  contig lengths of 24,205 bp. All the assembly data were deposited in the DDBJ/EMBL/GenBank nucleotide sequence database.

The draft genome of MSP20.1 was annotated using the RAST server (8), Glimmer 3 (9, 10), Genemark (11), tRNAScan-SE (12), and RNAmmer (13) for predicting subsystems, coding sequences (CDS), tRNAs, and rRNA genes, etc.

Using the different software tools, we identified 4,255 CDS in the draft genome of MSP20.1, with 66 RNA genes (64 tRNA and 2 rRNA genes) and 433 subsystems. Among the CDS, 2,474 are not in a subsystem (1,050 nonhypothetical, 1,424 hypothetical), whereas 1,781 CDS (nonhypothetical, 1,692; hypothetical, 89) are in a subsystem. RAST annotation also revealed the association of 110 genes in stress responses: 26 in osmotic stress (2 in osmoregulation, 20 in choline and betaine uptake and betaine biosynthesis, and 4 in ectoine biosynthesis and regulation), 46 in oxidative stress (11 in protection from reactive oxygen species [ROS], 24 in oxidative stress, 2 in glutathione:nonredox reactions, 6 in redox-dependent regulation, 2 in glutaredoxins, and 1 in the CoA disulfide thiol-disulfide redox system), 2 in cold shock, 14 in heat shock, 1 in detoxification, and 21 in no subcategory. In addition, 410 genes have been mapped to different pathways involved in the biosynthesis and degradation of amino acids and derivatives, including 105 in branched-chain amino acid pathways and 32 in the serine-glyoxylate cycle.

Further studies are under way to understand the mechanism(s) of the moderate halophilism of *Bacillus megaterium* strain MSP20.1 and its survival in salty settings.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [AVBB00000000](https://www.ncbi.nlm.nih.gov/nuccore/AVBB00000000). The version described in this paper is version AVBB01000000.

## ACKNOWLEDGMENTS

The work was carried out in the subproject “Exploring bacterial and archaeobacterial diversity in the Kutch ecoregion of Gujarat for agricultural and industrial applications” under the Application of Microorganisms in Agriculture and Allied Sectors (AMAAS) of the Indian Council of Agricultural Research (ICAR). We thank ICAR for funding through AMAAS.

We thank the directors of the Directorate of Groundnut Research, Junagadh, and NBAIM, Mau, for help during the course of this study. We also thank Macrogen, Inc., South Korea, and Sequencher Tech Pvt., Ltd., Ahmedabad, India, for the genomic services provided.

## REFERENCES

1. Pal KK, Dey R, Sherathia D, Dalsania T, Savsani K, Patel I, Thomas M, Ghorai S, Vanpariya S, Rupapara R, Acharya N, Rawal P, Joshi P, Sukhadiya B, Mandaliya M, Saxena AK. 2013. Draft genome sequence of *Salinibacillus aidingensis* strain MSP4, an obligate halophilic bacterium isolated from a salt crystallizer of the Rann of Kutch, India. *Genome Announc.* 1(4):e00253-13. <http://dx.doi.org/10.1128/genomeA.00253-13>.
2. Pal KK, Dey R, Thomas M, Sherathia D, Dalsania T, Patel I, Savsani K, Ghorai S, Vanpariya S, Sukhadiya B, Mandaliya M, Rupapara R, Rawal P, Saxena AK. 2013. Draft genome sequence of *Bacillus* sp. strain SB47, an obligate extreme halophile isolated from a salt pan of the Little Rann of Kutch, India. *Genome Announc.* 1(5):e00816-13. <http://dx.doi.org/10.1128/genomeA.00816-13>.
3. Dey R, Pal KK, Sherathia D, Dalsania T, Savsani K, Patel I, Thomas M, Ghorai S, Vanpariya S, Rupapara R, Rawal P, Sukhadiya B, Mandaliya M, Saxena AK. 2013. Draft genome sequence of *Bacillus* sp. strain NSP9.1, a moderately halophilic bacterium isolated from the salt marsh of the Great Rann of Kutch, India. *Genome Announc.* 1(5):e00835-13. <http://dx.doi.org/10.1128/genomeA.00835-13>.
4. Pal KK, Dey R, Thomas M, Sherathia D, Dalsania T, Patel I, Savsani K, Ghorai S, Vanpariya S, Sukhadiya B, Mandaliya M, Rupapara R, Rawal P. 2013. Draft genome sequence of the extremely halophilic *Bacillus* sp. strain SB49, isolated from a salt crystallizer pond of the Little Rann of Kutch, India. *Genome Announc.* 1(5):e00869-13. <http://dx.doi.org/10.1128/genomeA.00869-13>.
5. Dey R, Pal KK, Sherathia D, Dalsania T, Savsani K, Patel I, Sukhadiya B, Mandaliya M, Thomas M, Ghorai S, Vanpariya S, Rupapara R, Rawal P, Saxena AK. 2013. Draft genome sequence of *Bacillus* sp. strain NSP2.1, a nonhalophilic bacterium isolated from salt marsh of the Great Rann of Kutch, India. *Genome Announc.* 1(5):e00909-13. <http://dx.doi.org/10.1128/genomeA.00909-13>.
6. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, Li S, Yang H, Wang J, Wang J. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272. <http://dx.doi.org/10.1101/gr.097261.109>.
7. Richter DC, Schuster SC, Huson DH. 2007. OSLay: optimal syntenic layout of unfinished assemblies. *Bioinformatics* 23:1573–1579. <http://dx.doi.org/10.1093/bioinformatics/btm153>.
8. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
9. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679. <http://dx.doi.org/10.1093/bioinformatics/btm009>.
10. Salzberg SL, Delcher AL, Kasif S, White O. 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 26:544–548. <http://dx.doi.org/10.1093/nar/26.2.544>.
11. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29:2607–2618. <http://dx.doi.org/10.1093/nar/29.12.2607>.
12. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
13. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.