genomeA
Journals.ASM.org

# Draft Genome Sequence of *Bacillus* sp. Strain SB47, an Obligate Extreme Halophile Isolated from a Salt Pan of the Little Rann of Kutch, India

**Kamal Krishna Pal,[a] Rinku Dey,[a] Manesh Thomas,[a] Dharmesh Sherathia,[a] Trupti Dalsania,[a] Ilaxi Patel,[a] Kinjal Savsani,[a] Sucheta Ghorai,[a] Sejal Vanpariya,[a] Bhoomika Sukhadiya,[a] Mona Mandaliya,[a] Rupal Rupapara,[a] Priya Rawal,[a] Anil Kumar Saxena[b]**

Microbiology Section, Directorate of Groundnut Research, Junagadh, Gujarat, India[a]; Division of Microbiology, Indian Agricultural Research Institute, New Delhi, India[b]

**Here, we report the 4.46-Mbp draft genome sequence of *Bacillus* sp. strain SB47, an extreme halophile isolated from a salt pan of the Little Rann of Kutch, India. Exploring the genome of this organism will facilitate the understanding and isolation of the gene(s) involved in its extreme osmotolerance.**

**Received** 11 September 2013 **Accepted** 16 September 2013 **Published** 10 October 2013

**Citation** 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. doi: 10.1128/genomeA.00816-13.



Address correspondence to Kamal Krishna Pal, pal@nrcg.res.in.

*Bacillus* sp. strain SB47, an obligate extreme halophile and endospore-forming bacterium, was isolated from a salt pan of the Little Rann of Kutch, India. It grows optimally at a 15% NaCl (range, 5 to 35%) concentration in medium at 37°C and at pH 7.5. The genome of *Bacillus* sp. strain SB47 was sequenced to understand the mechanisms of its extreme osmotolerance and to isolate the relevant gene(s).

The whole genome of *Bacillus* sp. SB47 (G+C content of 45.50%) was sequenced using the Roche 454 Genome Sequencer (GS FLX) at Macrogen Inc., South Korea, through Sequencher Tech Pvt. Ltd., Ahmedabad, India, by both shotgun and mate-paired library sequencing. In shotgun sequencing, an average read length of 360 bp was generated from 603,934 reads of 217,996,405 bases. Sequencing of the mate-pair libraries gave 151,827 and 131,939 reads, respectively, with average read lengths of 475 bp and 455 bp, respectively.

*De novo* assembly was performed using the GS *de novo* Assembler version 2.6 (1) with approximately 76-fold coverage, and 10 scaffolds of 4,468,918 bp and 33 scaffold contigs of 4,464,274 bp with average lengths of 446,891 bp and 135,281 bp, respectively, were obtained. An $N_{50}$ scaffold length of 2,486,969 bp (4,808 bp and 2,486,969 bp for the smallest and largest scaffolds, respectively) was obtained. Similarly, an $N_{50}$ contig length of 231,087 bp (1,620 bp and 807,418 bp for the smallest and largest contigs, respectively) was obtained. All assembly data were deposited in the DDBJ/EMBL/GenBank nucleotide sequence database.

The draft genome sequence was annotated by the RAST server (2), Glimmer 3 (3, 4), GeneMark (5, 6), the KEGG database (7), tRNAScan-SE (8), RNAmmer (9), and Signal P4.1 (10).

Using the different softwares, we predicted 4,718 coding sequences (CDSs), with 3,901,977 bp in the CDSs. There were 74 RNA-encoding genes (68 tRNA, 6 rRNA) and 396 subsystems. Among the CDSs, 2,650 are not in a subsystem (1,047 nonhypothetical CDSs, 1,603 hypothetical CDSs), whereas 2,068 CDSs (1,929 nonhypothetical, 139 hypothetical) are in a subsystem. RAST annotation also revealed the association of 105 genes involved in stress responses in this organism: 10 in osmotic stress (1 in osmoregulation, 9 in choline and betaine uptake and betaine biosynthesis), 46 in oxidative stress (7 in protection from reactive oxygen species [ROS], 28 in oxidative stress, 1 in NADPH:quinine oxidoreductase 2, 1 in glutathione:nonredox reactions, 6 in redox-dependent regulation of nucleus processes, and 3 in glutaredoxins), 1 in cold shock, 16 in heat shock, 10 in detoxification, 1 in periplasmic stress, and 21 in no subcategory, with 237 signal peptides. Similarly, 2,186 CDSs were mapped to different biochemical pathways of KEGG (K00003 to K16706). The genes responsible for the production of different enzymes for the biosynthesis of valine, leucine, and isoleucine (map00290) and a number of genes involved in ABC transporters (map02010), including transporters for alkanesulfonate (SsuA, SsuC), glycine betaine/proline (ProX, ProW, ProV), osmoprotectants (OpuBC, OpuBB, OpuBA), and phosphate transporters (PstA, PstB, PstC, PstS), were also mapped. Similarly, genes for two-component systems (map02020), like those involved in the response to $K^+$ limitation and $K^+$ transport (KdpD, KdpA, KdpB, KpdC) and genes for salt stress degradative enzymes (DegS, DegU), have been mapped.

Deciphering the genome of this organism further will facilitate the understanding of obligate and extreme halophilism and the genes, biochemical pathways, and metabolites involved in osmotolerance.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. ATNR00000000. The version described in this paper is version ATNR01000000.

## ACKNOWLEDGMENTS

The work was carried out in the subproject "Diversity analysis of *Bacillus* and other pre-dominant genera in extreme environments and its utilization in agriculture" of the National Agricultural Innovation Project (NAIP) and the Application of Microorganisms in Agriculture and Allied

Sectors (AMAAS) of the Indian Council of Agricultural Research (ICAR). We thank ICAR for funding through NAIP and AMAAS.

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

## REFERENCES

1. **454 Life Sciences Corporation.** 2011. 454 sequencing system software manual, version 2.6. 454 Life Sciences Corporation, Branford, CT.
2. **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. doi:10.1186/1471-2164-9-75.
3. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23:**673–679.
4. **Salzberg SL, Delcher AL, Kasif S, White O.** 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. **26:** 544–548.
5. **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.
6. **Lukashin AV, Borodovsky M.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26:**1107–1115.
7. **Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M.** 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res. **32:** 277–280.
8. **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.
9. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
10. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8:**785–786.