

Draft Genome Sequence of the Obligate Halophilic *Bacillus* sp. Strain NSP22.2, Isolated from a Seasonal Salt Marsh of the Great Rann of Kutch, India

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

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

Here, we report the 4.0-Mbp draft genome of an obligate halophile, *Bacillus* sp. strain NSP22.2, isolated from a seasonal salt marsh of the Great Rann of Kutch, India. To understand the mechanism(s) of obligate halophilism and to isolate the relevant gene(s), the genome of *Bacillus* sp. NSP22.2 was sequenced.

Received 19 November 2013 Accepted 26 November 2013 Published 19 December 2013

Citation Dey R, Pal KK, Sherathia D, Vanpariya S, Patel I, Dalsania T, Savsani K, Sukhadiya B, Mandaliya M, Thomas M, Ghorai S, Rupapara R, Rawal P. 2013. Draft genome sequence of the obligate halophilic *Bacillus* sp. strain NSP22.2, isolated from a seasonal salt marsh of the Great Rann of Kutch, India. *Genome Announc.* 1(6):e01104-13. doi:10.1128/genomeA.01104-13.

Copyright © 2013 Dey et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](#).

Address correspondence to Rinku Dey, rinku@nrcg.res.in or rinku_dy@yahoo.co.in.

The genomes of a number of species of *Bacillus*, isolated from the extremely saline environments of the Rann of Kutch, India, have been sequenced recently to understand the mechanisms of osmotolerance (1–4). *Bacillus* sp. strain NSP22.2 (16S rRNA GenBank accession no. JF802172), isolated from a seasonal salt marsh of the Great Rann of Kutch, India, grows optimally at a concentration of 7.0% NaCl (range, 5 to 20%) in growth medium, at pH 7.0 and 37°C. The genome of *Bacillus* sp. strain NSP22.2 was sequenced to understand its obligate and moderate halophilism and to isolate the relevant gene(s).

The Illumina HiSeq 2000 platform was used to sequence the genome of NSP22.2 at Macrogen, Inc., South Korea, through Sequencher Tech Pvt. Ltd., Ahmedabad, India. Originally, a total of 66,480,406 paired-end reads of 6,714,521,006 bases were generated, with a length of 72 nucleotides. After filtering (criteria, >90% of bases with a quality value [QV] of >20 for both ends) and the removal of PCR duplicates, 54,215,856 reads of 5,475,801,456 bases were obtained.

The draft genome was *de novo* assembled by SOAPdenovo assembler version 1.05 (5), with genome coverage of about 1,200×. The long assembly, along with ordering of the contigs, was generated by OSLay (6). The draft assembly of NSP22.2 (G+C content of 43.99%) resulted in an approximate genome size of 4,000,934 bases, contained in 20 scaffolds (average scaffold, 200,047 bp; minimum, 1,155 bp; maximum, 1,400,974 bp) and 401 contigs (294 contigs of >500 bp) of 3,996,217 bp, with an average contig length of 9,965 bp (maximum, 80,648 bp; minimum, 84 bp). All the assembly data were deposited in the DDBJ/EMBL/GenBank nucleotide sequence database.

The draft genome sequence of NSP22.2 was further annotated using the RAST server (7), Glimmer 3 (8, 9), GeneMark (10), tRNAscan-SE (11), RNAmmer (12), and the KEGG database (13) for predicting subsystems, coding sequences (CDSs), tRNA and rRNA genes, KEGG pathways, and more.

Using the softwares, 4,251 CDSs were identified in the draft

genome sequence of NSP22.2, with 53 RNA genes (51 tRNA and 2 rRNA) and 416 subsystems. Among the CDSs, 2,604 are not in a subsystem (967 nonhypothetical, 1,637 hypothetical), whereas 1,647 CDSs (1,560 nonhypothetical, 87 hypothetical) are in subsystems. RAST annotation also predicted the involvement of 105 genes in stress responses, including 27 genes involved in osmotic stress (2 in osmoregulation, 21 in choline and betaine uptake and betaine biosynthesis, and 4 in ectoine biosynthesis and regulation), 38 in oxidative stress (9 in protection from reactive oxygen species [ROS], 20 in oxidative stress, 2 in glutathione:nonredox reactions, 5 in redox-dependent regulation, 2 in glutaredoxins), 2 in cold shock, 15 in heat shock, 1 in detoxification, and 22 in no subcategory. Also, 7 genes involved in potassium homeostasis, 10 in glycerol and glycerol-3-phosphate uptake and utilization, 8 in mannitol utilization, 314 in biosynthesis and utilization of amino acid and derivatives, and 447 in utilization of carbohydrates, as well as other genes, were identified. Moreover, 1,976 genes were mapped in different pathways of KEGG (K00003 to K16697).

Work is in progress to unravel the mechanisms of obligate halophilism and adaptation in *Bacillus* sp. NSP22.2 and to identify the gene(s) relevant to osmotolerance.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [AVCV00000000](#). The version described in this paper is version [AVCV01000000](#).

ACKNOWLEDGMENTS

The work was carried out in the subproject “Exploring bacterial and archaeobacterial diversity in Kutch eco-region 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 also thank 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. 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.
2. 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. doi:10.1128/genomeA.00835-13.
3. 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. doi:10.1128/genomeA.00869-13.
4. 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 the salt marsh of the Great Rann of Kutch, India. *Genome Announc.* 1(5):e00909-13. doi:10.1128/genomeA.00909-13.
5. 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.
6. Richter DC, Schuster SC, Huson DH. 2007. OSLay: optimal syntenic layout of unfinished assemblies. *Bioinformatics* 23:1573–1579.
7. 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.
8. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
9. Salzberg SL, Delcher AL, Kasif S, White O. 1998. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 26:544–548.
10. 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.
11. 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.
12. 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.
13. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32:277–280.