

Complete Genome Sequence and Methylome Analysis of *Bacillus* Strain X1

Alexey Fomenkov, Keith D. Lunnen, Zhenyu Zhu, Brian P. Anton, Geoffrey G. Wilson, Tamas Vincze, Richard J. Roberts

New England Biolabs, Ipswich, Massachusetts, USA

***Bacillus* strain X1 is the source strain for the restriction enzyme BstXI. Its complete sequence and full methylome was determined using single-molecule real-time (SMRT) sequencing.**

Received 30 December 2014 Accepted 5 January 2015 Published 19 February 2015

Citation Fomenkov A, Lunnen KD, Zhu Z, Anton BP, Wilson GG, Vincze T, Roberts RJ. 2015. Complete genome sequence and methylome analysis of *Bacillus* strain X1. *Genome Announc* 3(1):e01593-14. doi:10.1128/genomeA.01593-14.

Copyright © 2015 Fomenkov et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Richard J. Roberts, roberts@neb.com.

Bacillus strain X1 (originally called *Bacillus stearothermophilus* X1) was originally provided by Dr. Neil Welker, while he was at Northwestern University, and is now housed in the New England Biolabs culture collection (NEB278). It is the original source of the type II restriction enzyme (REase) BstXI, first isolated in 1981 and sold commercially in 1983 (1). This enzyme recognizes and cleaves the bipartite DNA sequence CCAN ↓ NNNNNTGG. The complete genome sequence for *Bacillus* strain X1 was determined using a Pacific Biosciences (PacBio, Menlo Park, CA) RSII instrument (2). Three single-molecule real-time (SMRT) cells were used, generating 1,100 Mb of sequence which was assembled into a closed circular genome of 3,422,674 bp (>300 fold coverage). Using the SMRTPortal 2.2 analysis platform, two symmetrically methylated sequence motifs were detected, CC^{6m}ANNNNNTGG, corresponding to the recognition sequence for M.BstXI, and G^{6m}ATC, which corresponded to the methylation target of the three BstXII restriction-modification system methylases (^{6m}A = N6-methyladenine). The genes responsible for these modifications were identified by amino acid sequence comparisons with previously characterized DNA methyltransferases (MTases) as described previously (3). HW35_9765 and HW35_09760 encode the modification and specificity subunits of a typical type I MTase which, characteristically, modify bipartite DNA sequences. Subcloning into *Escherichia coli* confirmed that these code for the M.BstXI MTase, and that a nearby gene, HW35_9755, codes for the BstXI REase. The complete system is shown schematically in REBASE (<http://rebase.neb.com/cgi-bin/seqget?BstXI>) (1). Restriction modification (R-M) systems composed of a type II REase partnered with type I MTase have been found previously (4). M.BstXI is unusual, however, because its specificity subunit contains two recognition domains yet recognizes a sequence that is symmetric overall, implying that both domains recognize the same, 5'-CCA-3', half-sequence. These domains display close similarity to each other and to the single recognition domain of M.XcmI, which methylates the related sequence, CC^{6m}ANNN NNNNNTGG. Also present in the genome is a more complicated R-M system comprising one REase gene (HW35_01470) and three

MTase genes (REBASE: <http://rebase.neb.com/cgi-bin/seqget?BstXII>). All four encode proteins with strong similarity to well-characterized enzymes recognizing GATC (5). Subcloning demonstrated that these genes are active, although only that encoding M2.BstXII resulted in complete modification. One additional gene, HW35_12570, encoded a protein (BstORFCP), with significant similarity to a family of methyl-dependent REases found in *Staphylococcus aureus* (e.g., SauUSI, which recognizes S^{5m}CNGS) (6).

Nucleotide sequence accession number. This complete genome sequence has been deposited in DDBJ/ENA/GenBank under the accession no. CP008855 (BioProject PRJNA253895).

ACKNOWLEDGMENTS

This research was supported by New England Biolabs and the Small Business Innovation Research Program (National Institute of General Medical Sciences) of the National Institutes of Health award R44GM105125 to R.J.R.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Roberts RJ. 1983. Restriction and modification enzymes and their recognition sequences. *Nucleic Acids Res* 11:r135–r167.
2. Korlach J, Bjornson KP, Chaudhuri BP, Cicero RL, Flusberg BA, Gray JJ, Holden D, Saxena R, Wegener J, Turner SW. 2010. Real-time DNA sequencing from single polymerase molecules. *Methods Enzymol* 472: 431–455. [http://dx.doi.org/10.1016/S0076-6879\(10\)72001-2](http://dx.doi.org/10.1016/S0076-6879(10)72001-2).
3. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, Fomenkov A, Turner SW, Korlach J, Roberts RJ. 2012. The methylomes of six bacteria. *Nucleic Acids Res* 40:11450–11462. <http://dx.doi.org/10.1093/nar/gks891>.
4. Loenen WA, Dryden DT, Raleigh EA, Wilson GG. 2014. Type I restriction enzymes and their relatives. *Nucleic Acids Res* 42:20–44. <http://dx.doi.org/10.1093/nar/gkt847>.
5. Ueno T, Ito H, Kimizuka F, Kotani H, Nakajima K. 1993. Gene structure and expression of the *MboI* restriction-modification system. *Nucl Acids Res* 21:2309–2313. <http://dx.doi.org/10.1093/nar/21.10.2309>.
6. Xu S-y, Corvaglia AR, Chan S-H, Zheng Y, Linder P. 2011. A type IV modification-dependent restriction enzyme SauUSI from *Staphylococcus aureus* subsp. *aureus* USA300. *Nucleic Acids Res* 39:5597–5610. <http://dx.doi.org/10.1093/nar/gkr098>.