

Complete Genome Sequence Analysis of *Bacillus subtilis* T30

Shuang-yong Xu,^a Matthew Boitano,^b Tyson A. Clark,^b Tamas Vincze,^a Alexey Fomenkov,^a Sanjay Kumar,^a Priscilla Hiu-Mei Too,^a Danila Gonchar,^c Sergey K. Degtyarev,^c Richard J. Roberts^a

New England Biolabs, Inc., Ipswich, Massachusetts, USA^a; Pacific Biosciences, Menlo Park, California, USA^b; SibEnzyme Ltd., Novosibirsk, Russia^c

The complete genome sequence of *Bacillus subtilis* T30 was determined by SMRT sequencing. The entire genome contains 4,138 predicted genes. The genome carries one intact prophage sequence (37.4 kb) similar to *Bacillus* phage SPBc2 and one incomplete prophage genome of 39.9 kb similar to *Bacillus* phage phi105.

Received 18 March 2015 Accepted 25 March 2015 Published 7 May 2015

Citation Xu S-Y, Boitano M, Clark TA, Vincze T, Fomenkov A, Kumar S, Too PH-M, Gonchar D, Degtyarev SK, Roberts RJ. 2015. Complete genome sequence analysis of *Bacillus subtilis* T30. *Genome Announc* 3(3):e00395-15. doi:10.1128/genomeA.00395-15.

Copyright © 2015 Xu 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/4.0/).

Address correspondence to Shuang-yong Xu, xus@neb.com, or Richard J. Roberts, roberts@neb.com.

Bacillus subtilis T30 is the source strain for the methylation-dependent restriction endonuclease (REase) *BisI* ($G^{5m}C \downarrow$ NGC). *BisI* belongs to the type IIM group of REases that cleave modified DNA (1, 2). *BisI* strain isolation, its morphological and physiological characterization, as well as the native *BisI* enzyme property were described previously (3). Here, we report the complete genome sequence of *B. subtilis* T30. Six SMRT cells worth of data from long-insert libraries of *B. subtilis* T30 genomic DNA were obtained. The sequence data were processed using HGAP and Quiver for *de novo* assembly (4). The assembled genome consisted of a single contig of 4.03 Mbp with 4,138 predicted genes (3,896 predicted coding sequences [CDSs]). The *B. subtilis* T30 genome sequence is very similar to that of *B. subtilis* subsp. *spizizenii* W23 (5), except that it contains regions of large repeats that impart difficulty in contig assembly from short reads created by other sequencing methods.

We analyzed the sequence for possible DNA methyltransferases (MTases) and endonucleases in the *B. subtilis* T30 genome by sequence homology analysis with known type I to IV restriction-modification (RM) system components listed in REBASE (1). In addition, by measuring the time-resolved kinetics of dT incorporation opposite to dA or $d^{m6}A$ by SMRT sequencing, it is possible to determine the methylation status of the template strand (6). SMRT analysis identified one active type I MTase that must be encoded by the single type I RM system in the genome (*hsdM*, *Bis30_13985*; *hsdS*, *Bis30_13990*), as evidenced in the methylated motif 5' $AC^{m6}AYN_7TGNG$ 3' (T indicates that the complementary A is modified). The half sites $AC^{m6}AY$ and $CNC^{m6}A$ are 94.7% and 94.5% modified, respectively, in the sequenced genome for self-protection. By amino acid sequence homology analysis with known DNA MTases, two putative C5 MTases were found in the *B. subtilis* T30 genome. The first, M.BisIII, was active and modified the site CCWGG (*Bis30_09930*). A second C5 MTase (*Bis30_20265*) adjacent to the *BisI* endonuclease was inactive when cloned in *Escherichia coli*. A prophage-encoded HNH endonuclease (*Bis30_20225*) was found to be active and conferred the DNA nicking specificity of 5' YG \downarrow GT 3' in Mg^{2+} buffer (the down arrow indicates the nicking strand as shown). *Bis30_20225* nicking specificity is also similar to N. ϕ Gamma (5' CG \downarrow GT 3') (7, 8). We next evaluated a few open

reading frames encoding putative endonucleases. Cell extracts of a putative PLD family endonuclease (*Bis30_09935*) or purified protein of one HNH endonuclease (*Bis30_16040*) were inactive in cleaving modified plasmid DNA (pBR322-*fhu4HIM*, $G^{5m}CNGC$, substrate for the native *BisI* endonuclease) or λ DNA. Thus, *Bis30_09935* and *Bis30_16040* were excluded as candidates for *BisI* endonuclease.

Nucleotide sequence accession number. The complete genome sequence has been deposited in DDBJ/ENA/GenBank under the accession number CP011051.

ACKNOWLEDGMENTS

We thank Joanna Bybee, Erbay Yibit, Siu-Hong Chan, Janos Posfai, Nick Guan, Mike Dalton, Rick Morgan, Bill Jack, Penghua Zhang, and Steven Salzberg (University of Maryland) for help with this project.

This work was partially supported by New England Biolabs, Inc. The genome sequencing, assembly, and N6mA modified site analysis were carried out at Pacific Biosciences.

REFERENCES

- Roberts RJ, Vincze T, Posfai J, Macelis D. 2010. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 38:D234–D236. <http://dx.doi.org/10.1093/nar/gkp874>.
- Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, Blumenthal RM, Degtyarev SK, Dryden DT, Dybvig K, Firman K, Gromova ES, Gumpert RI, Halford SE, Hattman S, Heitman J, Hornby DP, Janulaitis A, Jeltsch A, Josephsen J, Kiss A, Klaenhammer TR, Kobayashi I, Kong H, Kruger DH, Lacks S, Marinus MG, Miyahara M, Morgan RD, Murray NE, Nagaraja V, Piekarowicz A, Pingoud A, Raleigh E, Rao DN, Reich N, Repin VE, Selker EU, Shaw PC, Stein DC, Stoddard BL, Szybalski W, Trautner TA, Van Etten JL, Vitor JM, Wilson GG, Xu SY. 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res* 31:1805–1812. <http://dx.doi.org/10.1093/nar/gkg274>.
- Chmuzh EV, Kashirina JG, Tomilova JE, Mezentzeva NV, Dedkov VS, Gonchar DA, Abdurashitov MA, Degtyarev SK. 2005. Restriction endonuclease *bisI* from *Bacillus subtilis* T30 recognizes methylated sequence 5'-G(m5C) \downarrow NGC-3'. *Biotechnologia (Russia)* 3:22–26.
- 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. <http://dx.doi.org/10.1038/nmeth.2474>.
- Zeigler DR. 2011. The genome sequence of *Bacillus subtilis* subsp. *spizizenii* W23: insights into speciation within the *B. subtilis* complex and into the

- history of *B. subtilis* genetics. *Microbiology* 157:2033–2041. <http://dx.doi.org/10.1099/mic.0.048520-0>.
6. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 7:461–465. <http://dx.doi.org/10.1038/nmeth.1459>.
 7. Xu SY, Gupta YK. 2013. Natural zinc ribbon HNH endonucleases and engineered zinc finger nicking endonuclease. *Nucleic Acids Res* 41: 378–390. <http://dx.doi.org/10.1093/nar/gks1043>.
 8. Gutjahr A, Xu SY. 2014. Engineering nicking enzymes that preferentially nick 5-methylcytosine-modified DNA. *Nucleic Acids Res* 42:e77. <http://dx.doi.org/10.1093/nar/gku192>.