



Complete Genome Sequence and Methyloome Analysis of *Bacillus caldolyticus* NEB414

Alexey Fomenkov,^a Tamas Vincze,^a Fana Mersha,^a Richard J. Roberts^a

^aNew England Biolabs, Ipswich, Massachusetts, USA

ABSTRACT *Bacillus caldolyticus* NEB414 is the original source strain for the restriction enzyme BclI. Its complete sequence and full methylome were determined using single-molecule real-time sequencing.

The bacterial strain *Bacillus caldolyticus* NEB414 was originally isolated as an extreme thermophile organism and is now housed in the New England Biolabs culture collection. It is the original source of the type II restriction enzyme BclI, which recognizes and cleaves the DNA sequence T↓GATCA, producing 5' cohesive ends compatible for cloning DNA digested with Sau3AI (↓GATC), BamHI (G↓GATCC), or BgIII (A↓GATCT) (1).

Genomic DNA from a culture of *B. caldolyticus* NEB414 was purified using a modified Qiagen method, and the genome was sequenced using the Pacific Biosciences (PacBio) RS II sequencing platform. Briefly, SMRTbell libraries were constructed from a genomic DNA sample sheared to an average size of ~10 to 20 kb using the G-tubes protocol (Covaris, Woburn, MA, USA), end repaired, and ligated to hairpin adapters. Incompletely formed SMRTbell templates were digested with a combination of exonuclease III and exonuclease VII (New England Biolabs, Ipswich, MA, USA). Qualification and quantification of genomic DNA fragments and SMRTbell libraries were performed using a Qubit fluorimeter (Invitrogen, Eugene, OR, USA) and a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). One 12-kb SMRTbell library was prepared according to the modified PacBio sample preparation protocols, including additional separation with the BluePippin system, and sequenced using C4-P6 chemistry and 3 single-molecule real-time (SMRT) cells with a 240-min collection time. Sequencing reads were processed, mapped, and assembled by the Pacific Biosciences SMRT analysis pipeline using the HGAP3 protocol and polished using Quiver (2) to yield 943.9 Mb of sequence data, which was assembled into two contigs, a closed circular genome of 3,457,580 bp with 159.9-fold coverage and a plasmid, pBcl1, of 28,002 bp with 362.8-fold coverage. The assembled sequences were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) and have been deposited at DDBJ/EMBL/GenBank (accession numbers CP025074 and CP025075).

The advantage of the PacBio sequencing platform is its ability to detect the epigenetic state of sequenced DNA, which allows for the identification of modified nucleotides and their corresponding motifs. Epigenetic modification at each nucleotide position was measured as kinetic variations (KVs) in the nucleotide incorporation rates, and methylated motifs were deduced from the KV data (3–5). Four DNA methyltransferase recognition motifs were detected, each containing m6A modifications. The motifs were then matched with methyltransferase genes in the genome, and the results are shown in Table 1. They have also been deposited in REBASE (6).

Received 3 January 2018 Accepted 9 January 2018 Published 8 February 2018

Citation Fomenkov A, Vincze T, Mersha F, Roberts RJ. 2018. Complete genome sequence and methylome analysis of *Bacillus caldolyticus* NEB414. *Genome Announc* 6:e01605-17. <https://doi.org/10.1128/genomeA.01605-17>.

Copyright © 2018 Fomenkov et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

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

TABLE 1 Summary of methyltransferases identified in *B. caldolyticus* NEB414

Motif ^a	Assigned name	Methylation type	Restriction modification type
TG A TCA	M.Bcl	m6A	II alpha
CG A NNNNNNNT A RC	M.BclII	m6A	I gamma
G C CA T	M.BclIII	m6A	III beta
G A G A G	M.BclIV	m6A	III beta

^aModified bases are shown in bold; they are underlined when on the complementary strand.

Accession number(s). The complete genome and pBclI plasmid sequences of the *B. caldolyticus* isolate reported here have been deposited in GenBank under the accession numbers [CP025074](#) and [CP025075](#).

ACKNOWLEDGMENTS

This project was supported by the Small Business Innovation Research Program (NIGMS) of the National Institutes of Health under award R44GM105125 to R.J.R.

A.F., T.V., F.M., and R.J.R. work for the New England Biolabs company, which sells research reagents, including restriction enzymes and DNA methylases, to the scientific community.

REFERENCES

1. Bingham AHA, Atkinson T, Sciaky D, Roberts RJ. 1978. A specific endonuclease from *Bacillus caldolyticus*. *Nucleic Acids Res* 5:3457–3467. <https://doi.org/10.1093/nar/5.10.3457>.
2. 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. <https://doi.org/10.1038/nmeth.2474>.
3. 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. <https://doi.org/10.1038/nmeth.1459>.
4. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts RJ, Korlach J. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res* 40:e29. <https://doi.org/10.1093/nar/gkr1146>.
5. Korlach J, Turner SW. 2012. Going beyond five bases in DNA sequencing. *Curr Opin Struct Biol* 22:251–261. <https://doi.org/10.1016/j.sbi.2012.04.002>.
6. Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res* 43:D298–D299. <https://doi.org/10.1093/nar/gku1046>.