



Complete Genome Sequence and Methyloome Analysis of *Deinococcus wulumuqiensis* 479

Alexey Fomenkov,^a Yvette Luyten,^a Tamas Vincze,^a Brian P. Anton,^a Richard J. Roberts,^a Richard D. Morgan^a

^aNew England Biolabs Inc., Ipswich, Massachusetts, USA

ABSTRACT *Deinococcus wulumuqiensis* 479 (formerly known as *Deinococcus radiodurans* 479) is the original source strain for the restriction enzyme DrdI. Its complete sequence and full methylome were determined using Pacific Biosciences single-molecule real-time (SMRT) sequencing.

Deinococcus wulumuqiensis 479 was isolated in 1988 as part of a screening program for restriction enzymes having new specificities and is curated in the New England Biolabs culture collection (NEB 479). It is the original source of the prototype type II restriction enzyme DrdI that recognizes and cleaves the DNA sequence 5'-GACNNNN ↓ NNGTC-3' (1). The genus *Deinococcus* is well known for extreme resistance to ionizing radiation, desiccation, and DNA-damaging chemicals (2).

Genomic DNA from an overnight L broth (3) liquid culture of *Deinococcus wulumuqiensis* 479 was purified using a modified protocol from a blood and cell culture DNA kit (Qiagen, USA) and sequenced using the Pacific Biosciences (PacBio) RS II sequencing platform. Briefly, SMRTbell libraries were constructed from a genomic DNA sample sheared to ~10 to 20 kb using the g-tubes protocol (Covaris, Woburn, MA, USA), end repaired, and ligated to PacBio hairpin adapters. Incompletely formed SMRTbell templates and linear DNAs were digested with a combination of exonuclease III and exonuclease VII (New England Biolabs, Ipswich, MA, USA). DNA qualification and quantification were performed using the Qubit fluorimeter (Invitrogen, Eugene, OR) and 2100 Bioanalyzer (Agilent Technology, Santa Clara, CA). Two 15-kb SMRTbell libraries were prepared according to modified PacBio sample preparation protocols, including additional separation on BluePippin (Sage Science, Beverly, MA), originally sequenced using C2-P4 chemistry (7 SMRT cells, 120-min collection times) and later sequenced with C4-P6 chemistry (1 SMRT cell, 300-minute collection time). Sequencing reads (391,202 reads, mean subread length of 4,278 bp, and yield of 1.7 Gb for C2-P4; and 38,636 reads, mean subread length of 8,667 bp, and yield of 335 Mb for C4-P6) were assembled *de novo* using Hierarchical Genome Assembly Process (HGAP) assembly 1 version 2.1.1 (C2-P4 sequence) or HGAP assembly 3 version 2.3.0 (C4-P6 sequence) with default quality and read length parameters and polished using Quiver (4). Both assemblies gave 7 closed-circular genome elements with a G+C content of 65.62% (Table 1). The assembled sequence was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (5, 6).

One advantage of the PacBio sequencing platform is its ability to detect the epigenetic state of sequenced DNA (7–9). Seven DNA methyltransferase recognition motifs were detected by single-molecule real-time (SMRT) motif and modification analysis version 2.1.1 and 2.3.0, each containing m6A modifications. While PacBio methylome analysis does not reliably detect m5C methylation, genome-wide m5C methylation analysis has been determined separately (B. P. Anton, unpublished data) to identify m5C recognition motifs. All motifs were matched with the responsible meth-

Citation Fomenkov A, Luyten Y, Vincze T, Anton BP, Roberts RJ, Morgan RD. 2019. Complete genome sequence and methylome analysis of *Deinococcus wulumuqiensis* 479. *Microbiol Resour Announc* 8:e00066-19. <https://doi.org/10.1128/MRA.00066-19>.

Editor Vincent Bruno, University of Maryland School of Medicine

Copyright © 2019 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 Alexey Fomenkov, fomenkov@neb.com, or Richard D. Morgan, morgan@neb.com.

Received 16 January 2019

Accepted 20 February 2019

Published 21 March 2019

TABLE 1 Summary of genome elements, methyltransferase genes, and their motifs identified in *Deinococcus wulumuqiensis* 479

Genetic element	GenBank accession no.	Genome size (bp)	Genome coverage (×)	Methylase (RM system) name ^a	Recognition motif ^b	Methylation, RM type	
Chromosome	CP031158	2,658,929	423.73	RM.DrdVIII ^c	ARGAGC	6mA, IIG	
				RM.DrdV ^c	CATGNAC	6mA, IIG	
				DrdIX	GNGAYNNNNNCTC	6mA, I	
Plasmids	pDrdI	CP031163	668,667	536.84	M.DrdI ^c	GACNNNNNNGTC	6mA, II
					M.DrdORFCP	Not active	6mA, I
					M.DrdVII	YCGCGR	5mC II
	pDrdA	CP031159	311,711	511.45			
	pDrdB	CP031160	70,951	678.58			
	pDrdIV	CP031162	31,268	544.7	RM.DrdIV ^c	TACGAC	6mA, IIG
	pDrdII	CP031161	21,536	544.9	RM.DrdII ^c	GAACCA	6mA, IIG
pDrdVI	CP031164	15,512	243.93	M1-2.DrdVI ^c	GCAGCC	6mA, IIG	

^a RM, restriction modification.

^b Modified bases are in bold and bases opposite to them are bold and underlined.

^c Methyltransferase (MTase) genes cloned and expressed in *E. coli* strain ER2796 (11). DrdIII activity has been previously reported (1). Recent analysis of this strain did not reveal any traces of DrdIII activity nor the presence of a CGATCG modified motif. The strain may have lost a plasmid encoding the DrdIII gene, or the enzyme called DrdIII may have come from a contaminant.

yltransferases through cloning and expression in *Escherichia coli*, and the results are shown in Table 1 and have been deposited in REBASE (10).

Data availability. The complete genome sequence of *Deinococcus wulumuqiensis* 479 is available in GenBank under the accession numbers CP031158, CP031159, CP031160, CP031161, CP031162, CP031163, and CP031164. Original sequence reads have been deposited in NCBI under SRA accession number SRS4110296 and BioProject accession number PRJNA482107.

ACKNOWLEDGMENTS

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

A.F., Y.L., B.P.A., T.V., R.J.R., and R.D.M. work for New England Biolabs, a company that sells research reagents, including restriction enzymes and DNA methyltransferases, to the scientific community.

REFERENCES

- Polisson C, Morgan RD. 1989. DrdI, a unique restriction endonuclease from *Deinococcus radiodurans* which recognizes 5'GACN6GTC3'. *Nucleic Acids Res* 17:3316. <https://doi.org/10.1093/nar/17.8.3316>.
- Zahradka K, Slade D, Bailone A, Sommer S, Averbek D, Petranovic M, Lindner AB, Radman M. 2006. Reassembly of shattered chromosomes in *Deinococcus radiodurans*. *Nature* 443:569–573.
- Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 62:293–300.
- 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. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44: 6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>.
- 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>.
- 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>.
- 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>.
- 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>.
- Anton BP, Mongodin EF, Agrawal S, Fomenkov A, Byrd DR, Roberts RJ, Raleigh EA. 2015. Complete genome sequence of ER2796, a DNA methyltransferase-deficient strain of *Escherichia coli* K-12. *PLoS One* 10:e0127446. <https://doi.org/10.1371/journal.pone.0127446>.