



HAL
open science

Complete Genome and Methyloome Analysis of the Box-Shaped Halophilic Archaeon *Haloarcula sinaiensis* ATCC 33800

Alexey Fomenkov, Priya Dassarma, Sean P. Kennedy, Richard J. Roberts,
Shiladitya Dassarma

► **To cite this version:**

Alexey Fomenkov, Priya Dassarma, Sean P. Kennedy, Richard J. Roberts, Shiladitya Dassarma. Complete Genome and Methyloome Analysis of the Box-Shaped Halophilic Archaeon *Haloarcula sinaiensis* ATCC 33800. *Microbiology Resource Announcements*, American Society for Microbiology, 2021, 10 (34), pp.e0061921. 10.1128/MRA.00619-21 . pasteur-03689621

HAL Id: pasteur-03689621

<https://hal-pasteur.archives-ouvertes.fr/pasteur-03689621>

Submitted on 7 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License



Complete Genome and Methylome Analysis of the Box-Shaped Halophilic Archaeon *Haloarcula sinaiensis* ATCC 33800

 Alexey Fomenkov,^a Priya DasSarma,^b Sean P. Kennedy,^c Richard J. Roberts,^a  Shiladitya DasSarma^{b,c}

^aResearch Department, New England Biolabs, Ipswich, Massachusetts, USA

^bDepartment of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

^cDepartment of Computational Biology, Institut Pasteur, Paris, France

ABSTRACT The genome of halophilic archaeon *Haloarcula sinaiensis* ATCC 33800 was sequenced and assembled and comprises seven replicons. Four m6A and one m4C modified motifs and their responsible methyltransferase genes have been identified in the genome by single-molecule real-time (SMRT) sequencing and bioinformatic analysis.

Haloarcula *sinaiensis* ATCC 33800 is a halophilic archaeon isolated from brine samples collected at the Red Sea Sabkha Gavish (1). The cells are flat and angular, including irregular rectangles, triangles, and squares. *H. sinaiensis* is extremely halophilic with a requirement for more than 2 M NaCl for growth and 3 to 4 M NaCl for optimal growth (2–4).

A Pacific Biosciences RS II instrument and SMRTAnalysis pipeline were used for sequencing, contig assembly, and modification analysis as in reference 5. The *H. sinaiensis* ATCC 33800 strain was a gift from ATCC. Briefly, cells were grown in 500 ml ATCC 1230 *Haloarcula* medium and trace elements (3.50 mg/ml FeSO₄ · 7H₂O, 0.88 mg/ml ZnSO₄ · 7H₂O, 0.66 mg/ml MnSO₄ · H₂O, and 0.02 mg/ml CuSO₄ · 5H₂O [pH 7.4]) diluted 2,000-fold at 37°C for 36 h and then centrifuged to recover cells. Pellets were resuspended in 100 ml of a 25% NaCl salt solution and lysed by osmotic shock by addition of a 10 mM EDTA solution until visual clearing occurred. High-molecular-weight DNA was harvested. The archaeal lysate was extracted three times with an equal volume of phenol saturated with Tris-HCl (pH 8) followed by dialysis and treatment with RNase (6).

A total of 10 μg of genomic DNA (gDNA) was used to prepare single-molecule real-time (SMRT) libraries. 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 (NEBnext end repair, E6050; New England Biolabs [NEB], Ipswich, MA, USA), and ligated (Quick Ligation kit, M22005; NEB) 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). DNA qualification and quantification were performed using the Qubit fluorimeter (Invitrogen, Eugene, OR) and 2100 Bioanalyzer (Agilent Technology, Santa Clara, CA).

Three 6-, 10-, and 13-kb SMRTbell libraries were prepared according to the PacBio sample preparation protocol (<https://www.pacb.com/wp-content/uploads/2015/09/Guide-Pacific-Biosciences-Template-Preparation-and-Sequencing.pdf>), including additional separation on a BluePippin system (Sage Science, Beverly, MA) and sequencing with C4-P6 chemistry, using 9 SMRT cells, as follows: 5 with non-size selected (6 and 10 kb) and 4 with size selected 13-kb libraries with a 360-minute collection time for each library. A total of 3,048 Gb sequencing data in 274,386 polymerase reads with mean subread lengths of 4,980 bp and *N*₅₀ subread length of 7,635 bp were obtained (~426× coverage) and *de novo* assembled using HGAP_Assembly.3 version 2.3.0 with default quality and read length parameters and 3 times polished using Quiver (7). The polished assemblies generated seven closed circular contigs (Table 1). The assembled sequences were annotated using the NCBI

Citation Fomenkov A, DasSarma P, Kennedy SP, Roberts RJ, DasSarma S. 2021. Complete genome and methylome analysis of the box-shaped halophilic archaeon *Haloarcula sinaiensis* ATCC 33800. Microbiol Resour Announc 10:e00619-21. <https://doi.org/10.1128/MRA.00619-21>.

Editor Kenneth M. Stedman, Portland State University

Copyright © 2021 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 Shiladitya DasSarma, sdassarma@som.umaryland.edu.

Received 17 June 2021

Accepted 30 July 2021

Published 26 August 2021

TABLE 1 Summary of genome elements, DNA methyltransferase genes, and motifs identified in *Haloarcula sinaiensis* ATCC 33800

Genetic element	Accession no.	Genome size (bp)	Genome coverage (×)	Methylase (RM ^a system) name	Recognition motif ^b	Methylation RM type
Chromosome	CP073366	3,093,724	426.63	M.Hsi33800I M.Hsi33800IV	CTAG CTCA(N6) TCG	II I
Hsi_p540	CP073368	540,110	587.39	M.Hsi33800II Hsi33800V	GATC CTCCAG	II II
Hsi_p388	CP073369	387,656	402.79			
Hsi_p204	CP073371	204,099	816.16			
Hsi_p139	CP073372	139,649	633.55			
Hsi_117	CP073367	116,596	568.55	M.Hsi33800III	CAG(N6) TCGC	I
Hsi_p55	CP073370	55,294	1,490.08			

^aRM, restriction-modification.

^bModified bases and the bases opposite to them are in bold and underlined, respectively.

Prokaryotic Genomes Annotation Pipeline (PGAP) (8, 9). Previously, the *H. sinaiensis* ATCC 33800 genome sequence was deposited in GenBank as 84 contigs based on short read sequencing as [AOLR01000000.1](https://doi.org/10.1093/nar/gkx1068).

Four m6A and one m4C modified DNA motifs were detected by the SMRT motif and modification analysis version 2.3.0 (10–12). Scanning the genome using the Seqware program (13) predicted nine methyltransferase (MTase) genes in the genome. All five motifs and their corresponding MTase genes were predicted based on homology (14) to previously known MTases. The results are presented in Table 1 and have been deposited in REBASE (15).

Data availability. Sequences were deposited in the following NCBI databases under the indicated accession numbers: BioProject, [PRJNA412908](https://doi.org/10.1093/nar/gkx1068); BioSample, [SAMN18811338](https://doi.org/10.1093/nar/gkx1068); GenBank, [CP073366](https://doi.org/10.1093/nar/gkx1068) to [CP073372](https://doi.org/10.1093/nar/gkx1068); and SRA, [SRS8750153](https://doi.org/10.1093/nar/gkx1068).

ACKNOWLEDGMENTS

A.F. and R.J.R. work for New England BioLabs, a company that sells research reagents, including restriction enzymes and DNA methyltransferases, to the scientific community.

The DasSarma laboratory was supported by NASA grant 80NSSC17K0263 and NIH grant AI139808.

REFERENCES

- Javor B, Requadt C, Stoeckenius W. 1982. Box-shaped halophilic bacteria. *J Bacteriol* 151:1532–1542. <https://doi.org/10.1128/jb.151.3.1532-1542.1982>.
- DasSarma S, DasSarma P. 2017. Halophiles. In *Encyclopedia of life science*. John Wiley & Sons, Ltd, Hoboken, NJ. <https://doi.org/10.1002/9780470015902.a0000394.pub4>.
- DasSarma S, Schwieterman EW. 2021. Early evolution of purple retinal pigments on Earth and implications for exoplanet biosignatures. *Int J Astrobiol* 20:241–250. <https://doi.org/10.1017/S1473550418000423>.
- DasSarma S, DasSarma P. 2015. Halophiles and their enzymes: negativity put to good use. *Curr Opin Microbiol* 25:120–126. <https://doi.org/10.1016/j.mib.2015.05.009>.
- Fomenkov A, Grabovich MY, Dubinina G, Leshcheva N, Mikheeva N, Vincze T, Roberts RJ. 2020. Complete genome sequences and methylome analysis of two environmental *Spirochaetes*. *Microbiol Resour Announc* 9:e00236–20. <https://doi.org/10.1128/MRA.00236-20>.
- Sambrook J, Russell DW. 2001. *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 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>.
- 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, Gwladz 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>.
- Murray IA, Clark T, 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. <https://doi.org/10.1093/nar/gks891>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 3:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- 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>.