



# Methylomes of Two Extremely Halophilic *Archaea* Species, *Haloarcula marismortui* and *Haloferax mediterranei*

Shiladitya DasSarma,<sup>a</sup> Alexey Fomenkov,<sup>b</sup> Satyajit L. DasSarma,<sup>a</sup> Tamas Vincze,<sup>b</sup> Priya DasSarma,<sup>a</sup> Richard J. Roberts<sup>b</sup>

<sup>a</sup>Institute of Marine and Environmental Technology, Department of Microbiology and Immunology, University of Maryland—Baltimore, Baltimore, Maryland, USA

<sup>b</sup>New England Biolabs, Ipswich, Massachusetts, USA

**ABSTRACT** The genomes of two extremely halophilic *Archaea* species, *Haloarcula marismortui* and *Haloferax mediterranei*, were sequenced using single-molecule real-time sequencing. The ~4-Mbp genomes are GC rich with multiple large plasmids and two 4-methyl-cytosine patterns. Methyl transferases were incorporated into the Restriction Enzymes Database (REBASE), and gene annotation was incorporated into the Haloarchaeal Genomes Database (HaloWeb).

*Haloarcula marismortui* and *Haloferax mediterranei* are extremely halophilic microorganisms in the third domain of life forms, the *Archaea*, isolated from hypersaline environments, the Dead Sea and a Spanish saltern, respectively (1, 2). They are of significant interest to the biotechnology industry and the astrobiology community due to their survival in multiple extreme conditions, including saturated salinity, desiccation, high levels of solar radiation, and large numbers of toxic ions (3). We targeted these two halophilic *Archaea* (*Haloarchaea*) species for characterization of methylation patterns and methyl transferases utilizing whole-genome single-molecule real-time (SMRT) sequencing.

*Haloarcula marismortui* ATCC 43049 and *Haloferax mediterranei* ATCC 33500 were obtained from the American Type Culture Collection (ATCC) and grown according to the instructions of the ATCC, and nucleic acids were extracted using a published method (4). Sequencing was performed using the PacBio RS II platform with a SMRTbell sequencing library prepared from 3 g genomic DNA from each microbe and randomly sheared to 20 kb using the G-tube protocol (Covaris, Woburn, MA, USA). The libraries were sequenced using two SMRT cells each with P6-C4 chemistry and 360-min collection times. Sequencing reads were filtered (quality,  $\geq 0.80$ ; length,  $\geq 100$  bp) and assembled *de novo* (for *Haloarcula marismortui*, 121,886 sequencing reads with a mean length of 10,485 bp that gave a 1.278-Gb total sequence; for *Haloferax mediterranei*, 47,722 reads with a mean read length of 13,138 nucleotides yielding a 626-Mb total sequence) using HGAP\_Assembly.1 version 2.1.1 (P4-C2 sequence) and HGAP\_Assembly.3 version 2.3.0 (P6-C4 sequence), respectively, with default quality and read length parameters and polished using Quiver (5). Error correction and closure were performed using RS\_BridgeMapper.1 using the default parameters, and methylation patterns were determined using RS\_Modification\_and\_Motif\_Analysis.1 within SMRT Analysis using default settings (minimum modification QV, 30).

Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) build 3190 (6) and analyzed on the Haloarchaeal Genomes Database (HaloWeb version r1555192846) (7) using an installation of Lagan version 2.0 for comparison to the originally reported sequences (cf. [https://halo.umbc.edu/cgi-bin/haloweb/hma.pl?operation=pairwise\\_comparator](https://halo.umbc.edu/cgi-bin/haloweb/hma.pl?operation=pairwise_comparator) and [https://www.ncbi.nlm.nih.gov/genome/1084?genome\\_assembly\\_id=300522](https://www.ncbi.nlm.nih.gov/genome/1084?genome_assembly_id=300522); cf. [https://halo.umbc.edu/cgi-bin/haloweb/hme.pl?operation=pairwise\\_comparator](https://halo.umbc.edu/cgi-bin/haloweb/hme.pl?operation=pairwise_comparator), <https://www.ncbi.nlm.nih.gov/genome/11181>

**Citation** DasSarma S, Fomenkov A, DasSarma SL, Vincze T, DasSarma P, Roberts RJ. 2019. Methylomes of two extremely halophilic *Archaea* species, *Haloarcula marismortui* and *Haloferax mediterranei*. *Microbiol Resour Announc* 8:e00577-19. <https://doi.org/10.1128/MRA.00577-19>.

**Editor** Kenneth M. Stedman, Portland State University

**Copyright** © 2019 DasSarma 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](mailto:roberts@neb.com).

**Received** 21 May 2019

**Accepted** 6 June 2019

**Published** 3 July 2019

**TABLE 1** Methylated motifs in *Haloarcula marismortui* and *Haloferax mediterranei* detectable by SMRT sequencing

Species and motif <sup>a</sup>	Type	% motif detected <sup>b</sup>	Putative responsible MTase <sup>c</sup>
<i>H. marismortui</i>			
<b>CTAG</b>	4mC, II- $\beta$	89	M.HmaHMAI (ORF Hma_11876)
<u>TCGACGG</u>	4mC, II- $\alpha$	85	M.HmaHMAII (ORF Hma_6187)
<i>H. mediterranei</i>			
<b>CTAG</b>	4mC, II- $\beta$	97	M.Hme33500I (ORF HFX_760)
<u>HGCW<u>G</u>CK</u>	4mC, II- $\alpha$	83	M.Hme33500II (ORF HFX_3001)

<sup>a</sup>Locations of methylated bases are bold for the top strand and underlined for the bottom strand.

<sup>b</sup>% motif detected is a specific parameter of the SMRT motif and modification software and is dependent on a number of parameters, especially the genome coverage and nature of the modified base (m6A or m4C). For m4C detection, it almost never reaches 100% even when modification is 100%.

<sup>c</sup>Putative responsible MTase is the REBASE name for the methyltransferase responsible for the modification in the first column, with the open reading frame (ORF) number identified in parentheses. It was identified by sequence comparison with known methyltransferases of that specificity in the gold standard set in REBASE and the absence of other candidates in the genome.

?genome\_assembly\_id=227368, and [https://www.ncbi.nlm.nih.gov/genome/11181?genome\\_assembly\\_id=173551](https://www.ncbi.nlm.nih.gov/genome/11181?genome_assembly_id=173551)) (8). The genome sequences were very similar, with primarily single-nucleotide indels of G (or C) in short runs of G's (or C's) and a few insertion element indels. The *H. marismortui* genome was 4,184,929 bp long with a 62.4% GC content and was divided into a 4.2-Mbp chromosome and 6 plasmids of 33 to 411 kbp. The *H. mediterranei* genome was 3,907,473 bp with a 61.1% GC content and was divided into a 3.9-Mbp chromosome and 3 plasmids of 132 to 505 kbp. Like those of other *Haloarchaea* species, the *H. marismortui* and *H. mediterranei* genomes encode acidic proteomes with average protein pI values of 4.7 and 4.9, respectively (9). Both genomes encode all 799 conserved haloarchaeal orthologous groups (HOGs), including 77 signature-unique conserved HOGs characteristic of the family (10).

Both the *H. marismortui* and *H. mediterranei* genomes contain the methylated sequence m<sup>4</sup>CTAG. The corresponding type II- $\beta$  methyl transferases (MTase) are coded by *zim* (VNG1543) in the model species *Halobacterium* sp. strain NRC-1 and over 100 other *Haloarchaea* species and are recognized as HOG1170 (10–13). Despite the wide distribution of CTAG methylation in the family, the function of this methylation system is still not known (14).

The *H. marismortui* and *H. mediterranei* genomes each also contain a single type II- $\alpha$  MTase enzyme forming m4C (Table 1). For *H. marismortui*, the enzyme is coded on a 155-kbp plasmid (pNG600/pHMA155), while for *H. mediterranei*, the enzyme is coded on the large chromosome.

**Data availability.** The *H. marismortui* genome sequence has been deposited in GenBank with the accession numbers [CP039132](#), [CP039133](#), [CP039134](#), [CP039135](#), [CP039136](#), [CP039137](#), and [CP039138](#). The *H. mediterranei* genome sequence has been deposited in GenBank with the accession numbers [CP039139](#), [CP039140](#), [CP039141](#), and [CP039142](#). Both genomes are also available on HaloWeb (<https://halo.umbc.edu/>) and have been analyzed in the Restriction Enzymes Database (REBASE) (<http://rebase.neb.com/rebase/rebase.html>) (15). Raw data are available in the NCBI Sequence Read Archive with the accession numbers [SRR8914802](#), [SRR8914803](#), [SRR8985819](#), and [SRR8985820](#).

## ACKNOWLEDGMENTS

The DasSarma laboratory is supported by NASA Exobiology grants NNX15AM07G and NNH18ZDA001N-EXO.

We thank Brian Anton for helpful discussions.

A.F., T.V., 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.

## REFERENCES

1. Ginzburg M, Sachs L, Ginzburg BZ. 1970. Ion metabolism in a *Halobacterium*. I. Influence of age of culture on intracellular concentrations. *J Gen Physiol* 55:187–207. <https://doi.org/10.1085/jgp.55.2.187>.
2. Valera Rodriguez F, Ruiz-Berraquero F, Ramos-Cormenzana A. 1980. Isolation of extremely halophilic bacteria able to grow in defined inorganic media with single carbon sources. *Microbiology* 119:535–538. <https://doi.org/10.1099/00221287-119-2-535>.
3. DasSarma S, DasSarma P. 2017. Halophiles, encyclopedia of life science. In eLS. John Wiley & Sons Ltd., Chichester, United Kingdom. <https://doi.org/10.1002/9780470015902.a0000394.pub4>.
4. Ng WL, Yang CF, Halladay JT, Arora P, DasSarma S. 1995. Protocol 25. Isolation of genomic and plasmid DNAs from *Halobacterium halobium*, p 179–184. In DasSarma S, Fleischmann EM (ed), *Archaea, a laboratory manual: halophiles*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
5. 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>.
6. Tatusova T, DiCuccio M, Badretdin A, Chetvermin 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>.
7. DasSarma SL, Capes MD, DasSarma P, DasSarma S. 2010. HaloWeb: the Haloarchaeal Genomes Database. *Saline Systems* 6:12. <https://doi.org/10.1186/1746-1448-6-12>.
8. Brudno M, Do C, Cooper G, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13:721–731. <https://doi.org/10.1101/gr.926603>.
9. 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>.
10. Capes MD, DasSarma P, DasSarma S. 2012. The core and unique proteins of haloarchaea. *BMC Genomics* 13:39. <https://doi.org/10.1186/1471-2164-13-39>.
11. Capes MD, Coker JA, Gessler R, Grinblat-Huse V, DasSarma SL, Jacob CG, Kim J-M, DasSarma P, DasSarma S. 2011. The information transfer system of halophilic archaea. *Plasmid* 65:77–101. <https://doi.org/10.1016/j.plasmid.2010.11.005>.
12. DasSarma S. 1995. Restriction-modification systems for halophiles, p 257. In DasSarma S, Fleischmann EM (ed). *Archaea, a laboratory manual: halophiles*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
13. Fullmer MS, Ouellette M, Louyakis AS, Papke RT, Gogarten JP. 2019. The patchy distribution of restriction-modification system genes and the conservation of orphan methyltransferases in halobacteria. *Genes (Basel)* 10:233. <https://doi.org/10.3390/genes10030233>.
14. Ouellette M, Gogarten JP, Lajoie J, Makkay AM, Papke RT. 2018. Characterizing the DNA methyltransferases of *Haloferax volcanii* via bioinformatics, gene deletion, and SMRT sequencing. *Genes (Basel)* 9:129. <https://doi.org/10.3390/genes9030129>.
15. 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>.