Complete Genome Sequence of *Halorubrum ezzemoulense* Strain Fb21

Yutian Feng,a Artemis S. Louyakis,a Andrea M. Makkay,a Ray O. Guerrero,a R. Thane Papke,a J. Peter Gogartena

aDepartment of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, USA

**ABSTRACT** Isolated from Aran-Bidgol Lake in Iran, and reported here, *Halorubrum ezzemoulense* strain Fb21 represents the first complete genome from this archaeal species. Local recombination in this genome is in stark contrast to equidistant recombination events in bacteria. The genome’s GC bias, however, points to a genome architecture and origin that resemble those of a bacterium. Its availability, genome signatures, and frequent intragenomic recombination mean that Fb21 presents an attractive model organism for this species.

*Halorubrum ezzemoulense* is a globally distributed species (1) of the extremely halophilic archaea that commonly dominate alkaline lakes, salterns, and other hypersaline environments. *H. ezzemoulense* DSM 17463T was isolated in 2006 (2) from Ezzemoul sabkha in Algeria. Despite numerous studies of this species and many high-quality draft genome sequence deposits in the NCBI database, to date there has been no complete genome sequence of *H. ezzemoulense*. Here, we present the first complete genome sequence of *H. ezzemoulense* strain Fb21, cultivated from hypersaline Aran-Bidgol Lake in Iran.

Fb21 was sampled in November 2007 from the shallow brine water column (10 cm, >200 practical salinity units) of Aran-Bidgol Lake (34°31'25" N; 51°53'40" E; 2,400 km²) and isolated as previously described (3). Briefly, DNA from Fb21 was isolated from pure liquid cultures and grown in *Haloferax volcanii* medium containing yeast extract, peptone, casamino acids, and 18% saltwater at 37°C (1, 4) using a phenol-chloroform-isooamyl alcohol (25:24:1) extraction followed by ethanol precipitation to purify the DNA, as described in the *Halohandbook* (5). Sequencing was completed using both short- and long-read platforms. Libraries were constructed using the Nextera XT DNA library preparation kit; two libraries were prepared for sequencing using the MiSeq reagent kit v2 (paired end, 500 cycles), a third used the MiSeq reagent kit v2 (paired end, 300 cycles), and each library was sequenced on separate runs on the MiSeq platform (Illumina, San Diego, CA) at the UConn MARS Center for Open Research Resources and Equipment. High-molecular-weight DNA was similarly isolated, with an additional cleanup using Agencourt AMPure beads (Beckman Coulter, Indianapolis, IN), and sequenced using a PacBio RS II single-molecule real-time (SMRT) DNA sequencing system on one SMRT cell long-read sequencing run on the Pacific Biosciences platform at the Keck Biotechnology Resource Laboratory at Yale University. All of the following programs were executed with default parameters unless otherwise stated. Reads were quality trimmed with Sickle v1.33 (6) and hybrid assembled with high-quality long reads using the Unicycler v0.4.7 pipeline (7) on the bold and normal settings. Contigs from the two assemblies were investigated and reconciled in Bandage v0.8.1 (8) to create the final assembly of the Fb21 genome. The assembly was then polished using the short reads with Pilon v1.130 (9). Three replicons were assembled and circularized, including a chromosome (~3.1 Mbp, 68.46% GC content), a megaplasmid (~606 kbp, 57.36% GC content), and a plasmid (~57 kbp, 54.66% GC content). Total genome
coverage with only the short reads was 99.94%, which was increased to 100% with the addition of long reads. The average sequencing depth across the genome with the short reads was 136× (standard deviation [SD] = 54.8), while the PacBio read average depth was 41× (SD = 17). Fb21 was previously determined to be a strain of *H. ezzemoulense* via 16S rRNA sequencing (1) and halobacterial lineage marker phylogenies in CheckM v1.0.12 (10). In addition, genome synteny between Fb21 and its closest neighbor, *Halorubrum lacusprofundi* ATCC 49239, was analyzed using nucmer v3.0 (11) and plotted (Fig. 1a). Numerous recombination events are visible in the plot. None of

![Graph](mra.asm.org.png)

**Fig 1** Analyses of the *Halorubrum ezzemoulense* Fb21 chromosome. (a) Nucmer comparison of *H. ezzemoulense* Fb21 (x axis) to *H. lacusprofundi* ATCC 9239 (y axis). Plus-strand best hits are shown in purple, while minus-strand best hits are shown in blue. Axis units are the chromosomal base positions. (b) The orange line represents cumulative C occurrences minus G occurrences (C − G) throughout the Fb21 chromosome, starting at the predicted main origin of replication. Values are calculated by subtracting the number of Gs encountered from the number of Cs at each chromosomal location. The blue line represents cumulative C − G strand bias in a 500,000-bp rolling window, calculated by the same method.
the detected recombination events are equidistant to the inferred origin of replication, unlike those found in bacterial genomes (12).

Archaeal mode in Prokka v1.13 (13) and the NCBI Prokaryotic Genome Annotation Pipeline (14) were used to predict and annotate the coding sequences (CDS) in the complete genome of Fb21. Of the 3,661 CDS identified, 3,443 were predicted to be proteins and 78 were predicted tRNAs, rRNAs, and noncoding (ncRNAs), while the remaining 218 were identified as pseudogenes. Aragorn v1.2.38 (15) and TrNAScan-SE v2.0 (16) were used to annotate 70 tRNAs, 6 rRNAs (2 identical operons), and 2 ncRNAs, all of which are on the chromosome. Using the orc1/cdc6 gene (17) and read mapping, we predicted seven potential origins of replication on the chromosome, three on the megaplasmid, and one on the plasmid. The most likely origin of replication determined by similarity and GC strand bias (Fig. 1b) was rotated to the beginning of the chromosome. The cumulative and rolling window GC signatures indicate a bacterium-like genome architecture, which is unique to date among the *Halobacteria*.

We found that the chromosome of Fb21 contains a host of genes related to quorum sensing, rhodopsins, and chemotaxis, in addition to all necessary housekeeping genes. The megaplasmid contains two pseudogenes and two genes encoding restriction endonucleases, over fifty transposases, and an additional ribonucleotide reductase homolog. The genome shows evidence of local recombination events that are atypical of bacterial genomes. However, the GC strand bias profile of this archaen resembles that of a bacterial genome. The Fb21 complete genome, in combination with the many other *H. ezzemoulense* draft genomes, makes this strain an excellent target for studies of within-species and within-population gene flow.

**Data availability.** The complete genome sequences and reports have been deposited in GenBank under accession numbers CP034940, CP034941, and CP034942 within BioProject accession number PRJNA513349 (assembly accession number GCF_000126515). The short and long reads are also available under the same BioProject accession number.

**ACKNOWLEDGMENTS**

This work was supported through a grant from the National Science Foundation (NSF/MCB 1716046) within the BSF-NSF joint research program.

We thank Matthew Ouellette for the growing and sequencing of Fb21.

**REFERENCES**

1. Corral P, de la Haba RR, Infante-Dominguez C, Sánchez-Porro C, Amoozegar MA, Papke RT, Ventosa A. 2018. *Halorubrum chaoi*viator Mancinelli et al. 2009 is a later, heterotypic synonym of *Halorubrum ezzemoulense* Kharroub et al. 2006. 6. Emended description of *Halorubrum ezzemoulense* Kharroub et al. 2006. Int J Syst Evol Microbiol 68:3657–3665. https://doi.org/10.1099/ijsem.0.033005.
2. Kharroub K, Quesada T, Ferrer R, Fuentes S, Aguilera M, Boulahrouf A, Ramos-Cormenzana A, Monteoliva-Sánchez M. 2006. *Halorubrum ezzemoulense* sp. nov., a halophilic archaean isolated from Ezzemoul sabkha, Algeria. Int J Syst Evol Microbiol 56:1583–1588. https://doi.org/10.1099/ijsem.0.064272-0.
3. de la Haba RR, Corral P, Sánchez-Porro C, Infante-Dominguez C, Makkay A, Amoozegar M, Ventosa A, Papke RT. 2018. Genotypic and lipid analyses of strains from the archaeal genus *Halorubrum* reveal insights into their taxonomy, divergence, and population structure. Front Microbiol 9:512. https://doi.org/10.3389/fmicb.2018.00512.
4. Allers T, Nigo H-P, Mvevarech M, Lloyd RG. 2004. Development of additional selectable markers for the halophilic archaean *Halofex dis* from the 16S and rtpA genes. Appl Environ Microbiol 70:943–953. https://doi.org/10.1128/AEM.70.2.943-953.2004.
5. Dyall-Smith M. 2009. *Halobacteria* handbook: protocols for halarchaeal genetics, v.7.2. Haloarchaeal Genetics Laboratory. https://haloarchaea.com/halohandbook/
6. Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). https://github.com/najoshi/sickle.
7. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Uncycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
8. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
9. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long reads. Genome Biol 18:176. https://doi.org/10.1186/s13059-017-1235-z.
10. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. Pilon: an integrated pipeline for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
11. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1103/gr.180702.114.
12. Kurtz S, Phillippy A, Delcher AL, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
13. Tillier ER, Collins RA. 2000. Genome rearrangement by replication-directed translocation. Nat Genet 26:195–197. https://doi.org/10.1038/79918.
14. Seemann T, 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
15. 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.

15. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.

16. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964. https://doi.org/10.1093/nar/25.5.955.

17. Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T. 2007. Genetic and physical mapping of DNA replication origins in Haloferax volcanii. PLoS Genet 3:e77. https://doi.org/10.1371/journal.pgen.0030077.