Draft Genome Sequence of *Haloferax* sp. Strain ATB1, Isolated from a Semi-Arid Region in the Brazilian Caatinga

Wendel de Oliveira Castro,[a] Adriana Maria Torres-Ballesteros,[b] Cristina Rossi Nakayama,[c] Itamar Soares Melo,[d] Vivian Helena Pellizari,[b] Artur Silva,[a] Rommel Thiago Jucá Ramos[a]

Institute of Biological Sciences, Federal University of Pará, Belém, PA, Brazil[a]; Laboratory of Ecology of Micro-organisms, IO, University of São Paulo, São Paulo, Brazil[b]; Institute of Environmental, Chemical and Pharmaceutical Sciences, Federal University of São Paulo, Diadema, Brazil[c]; Laboratory of Environmental Microbiology, EMBRAPA Environment, Jaguariúna, São Paulo, Brazil[d]

Organisms in the *Haloferax* genus are extreme halophiles that grow in environments with pH values between 4 and 12, and temperatures between 0°C and 60°C. In the present study, a draft of the first *Haloferax* sp. strain ATB1 genome isolated from the region of Cariri (in Paraíba State, Brazil) is presented.

Address correspondence to Rommel Thiago Jucá Ramos, rommelramos@ufpa.br.

Organisms in the *Haloferax* genus, which belong to the Archaea domain, characteristically inhabit extreme environments. In particular, these organisms can grow in locations with salinities ranging from 10% to 37%, pH values between 4 and 12, and temperatures between 0°C and 60°C (1). For instance, *H. chudinovii* grows at a pH of 6.8, 42°C, and 16% salinity (2). Halophilic species, which are resistant to dehydration, exhibit low levels of cellular macromolecule denaturation and high thermal stability (3). Moreover, these species are resistant to ultraviolet (UV) light because they contain light-driven proton pumps and produce high levels of carotenoids (4). Thus, halophilic organisms can be used as models to help understand the fundamental limits of life (5) and to identify proteins with biotechnological potential.

The *Haloferax* sp. ATB1 specimen examined in the present study was isolated from the soil of the Cariri complex (in Paraíba, Brazil), a portion of the Caatinga biome that is affected by the desertification processes. The isolate was cultured at a pH of 7.4, temperatures ranging from 37°C to 50°C, and a salinity of approximately 25%. The genome was sequenced with the Ion Torrent PGM platform using a mate-paired library (6). A crude data quality analysis was performed using the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The data was submitted to quality filter to remove the reads with quality values less than Phred 20.

This approach resulted in 8,480.874 reads, which corresponded to a genomic coverage of 224× based on the size of the reference genome, *H. volcanni* DS2 (4.1 Mb, including the main chromosome [CP001956] and the plasmids pHV1 [CP001957], pHV2 [CP001954], pHV3 [CP001953], and pHV4 [CP001955]).

A hybrid approach that utilized the four assemblers MIRA (7), CLC Genomics Workbench (http://www.clcbio.com/), SeqMan (http://www.dnastar.com), and SPAdes (8) was adopted for assembly; these assemblers produced 20,586, 14,800, 2,455, and 10,475 contigs, respectively. These results were integrated to generate a single set of sequences. This sequence set was subsequently processed using the CISA program (9) to finalize gaps, producing a total of only 120 contigs with an N_{50} of 36,538 that totaled 4.1 Mb.

The contigs were subjected to an automatic annotation process using the RAST software (10). This process identified 4,517 coding sequences (CDSs) and 33 RNAs, which had a G+C content of 61%. The examined *Haloferax* sp. ATB1 genome contains 4,223,705 bp.

Nucleotide sequence accession number. The genomic sequence obtained in this study has been deposited in the DDBJ/EMBL/GenBank database under accession number JPES00000000.

REFERENCES

1. **Oren A, Arahal DR, Ventosa A.** 2009. Emended descriptions of genera of the family Halobacteriaceae. Int. J. Syst. Evol. Microbiol. **59**(Pt 3):637–642. http://dx.doi.org/10.1099/ijs.0.008904-0.

2. **Saralov AI, Baslerov RV, Kuznetsov BB.** 2013. Haloferax chudinovii sp. nov., a halophilic archaeon from Permian potassium salt deposits. Extremophiles **17**:499–504. http://dx.doi.org/10.1007/s00792-013-0534-8.

3. **Tehei M, Franzetti B, Maurel M-C, Vergne J, Hountondji C, Zaccai G.** 2002. The search for traces of life: the protective effect of salt on biological macromolecules. Extremophiles **6**(5):427–30. http://dx.doi.org/10.1007/s00792-002-0275-6.

4. **Naziri D, Hamidi M, Hassanzadeh S, Tarhriz V, Zanjani BM, Nazemyieh H, Hejazi MA, Hejazi MS.** 2014. Analysis of carotenoid production by Halorubrum sp. TBZ126; an extremely halophilic archeon from Urmia Lake. Adv. Pharm. Bull. **4**(1):61–67. http://dx.doi.org/10.5681/apb.2014.010.

5. **DasSarma S.** 2006. Extreme halophiles are models for astrobiology. Microbe Mag. **1:**120–126.

6. **Ramos RTJ, Carneiro AR, de Castro Soares S, Barbosa S, Varuzza L, Orabona G, Tauch A, Azevedo V, Schneider MP, Silva A.** 2013. High efficiency application of a mate-paired library from next-generation sequencing to postlight sequencing: Corynebacterium pseudotuberculosis as a case study for microbial de novo genome assembly. J. Microbiol. Methods **95:**441–447. http://dx.doi.org/10.1016/j.mimet.2013.06.006.

7. **Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Wetter T.** 2004. Using the miraEST Assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. **14:** 1147–1159. http://dx.doi.org/10.1101/gr.1917404.

8. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

9. **Lin S-H, Liao Y-C.** 2013. CISA: contig integrator for sequence assembly of bacterial genomes. PLoS One **8**(3):e60843. http://dx.doi.org/10.1371/journal.pone.0060843.

10. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9:**75. http://dx.doi.org/10.1186/1471-2164-9-75.