

## GENOME ANNOUNCEMENTS

### Complete Genome Sequence of *Haloarcula hispanica*, a Model Haloarchaeon for Studying Genetics, Metabolism, and Virus-Host Interaction

Hailong Liu,<sup>1</sup> Zhenfang Wu,<sup>1</sup> Ming Li,<sup>1</sup> Fan Zhang,<sup>1</sup> Huajun Zheng,<sup>2</sup> Jing Han,<sup>1</sup>  
Jingfang Liu,<sup>1</sup> Jian Zhou,<sup>1</sup> Shengyue Wang,<sup>2</sup> and Hua Xiang<sup>1\*</sup>

State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China,<sup>1</sup> and Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China<sup>2</sup>

Received 5 August 2011/Accepted 17 August 2011

***Haloarcula hispanica* is an extremely halophilic archaeon that has an unusually low restriction barrier and is therefore significant for studying archaeal genetics, metabolism, and virus-host interactions. Here we report the complete genome sequence (3,890,005 bp) of *H. hispanica* strain CGMCC 1.2049, consisting of two chromosomes and one megaplasmid.**

*Haloarcula hispanica* was first isolated from a solar saltern in Spain (8). It is exceptional among members of the *Halobacteriaceae* family in that it exhibits unusually low restriction activity and is therefore significant as one of the most tractable haloarchaea for archaeal studies (2), as well being useful as a vital host for investigations of viruses in hypersaline ecosystems (13). A transformation method for *H. hispanica* was established 20 years ago (2, 3), whereas the high-efficient knockout system for its genome manipulation was developed recently (11). In particular, *H. hispanica* has shown the ability to accumulate polyhydroxyalkanoate (PHA) when carbon sources are in excess. Although a number of key genes (*phaEC* and *phaB*) involved in PHA biosynthesis in *H. hispanica* have been identified (6, 7), the entire PHA biosynthetic and regulatory networks in haloarchaea remain to be established. Thus, *H. hispanica* would be a good haloarchaeon model organism after the genome sequence is made available. Here, we announce the complete genome sequence of *H. hispanica* CGMCC 1.2049 (= ATCC 33960), which was obtained from the China General Microbiological Culture Collection Center (CGMCC).

The genome sequence was determined mainly by the use of massively parallel pyrosequencing technology and a Roche 454 GS FLX sequencer. A total of 506,320 reads, counting up to 124,501,912 bases (32-fold coverage of the genome), were assembled using *GS de novo* Assembler software, which generated 29 contigs ranging from 500 to 916,380 bp, with bases having quality scores of 40 and above. Relationships of the contigs were determined by multiplex PCR (14). Gaps were filled by sequencing the PCR products and the

pSMART (Lucigen) shotgun library using ABI 3730xl capillary sequencers. Finally, the genome was assembled using Phred, Phrap, and Consed software (<http://www.phrap.org/phredphrapconsed.html>), and low-quality regions of the genome were resequenced. Putative coding sequences (CDSs) were identified by ZCURVE (5) and Glimmer 3 (4) software, and peptides shorter than 30 amino acids were eliminated. tRNA genes were predicted by tRNAScan-SE software (12) and rRNA genes were identified by RNAmmer 1.2 (10). Functional annotation of CDSs was performed by searching the NCBI nonredundant protein database and the Kyoto Encyclopedia of Genes and Genomes (KEGG) protein database (9) using BLASTP (1).

The *H. hispanica* genome consists of main chromosome I (2,995,271 bp) (G+C content, 63.7%; coding density, 88.1%), minichromosome II (488,918 bp) (G+C content, 57.0%; coding density, 80.3%), and megaplasmid pHH400 (405,816 bp) (G+C content, 59.9%; coding density, 84.4%). The genome encodes 3,859 predicted proteins, 48 tRNAs, and 3 copies of 16S-23S-5S rRNA on chromosomes (1 copy on chromosome I) plus 1 copy of 5S rRNA on pHH400. Compared to the genome of *H. marismortui*, another sequenced *Haloarcula* with two chromosomes and seven plasmids, up to 80% of predicted proteins of *H. hispanica* have counterparts in *H. marismortui*, whereas ~700 predicted proteins are only present in *H. hispanica*. Interestingly, although their main chromosomes are homologous and their megaplasmids are homologous, their minichromosomes are completely different. Comparative analysis performed with the genome of *H. marismortui* and other sequenced haloarchaeal genomes may provide further insight into the genetics, metabolism, origin, and evolution of the extremely halophilic archaea.

**Nucleotide sequence accession numbers.** The genome sequence of *H. hispanica* CGMCC 1.2049 was deposited in NCBI

\* Corresponding author. Mailing address: State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, No.1 West Beichen Road, Chaoyang District, Beijing 100101, China. Phone and fax: 86-10-64807472. E-mail: xiangh@im.ac.cn.

GenBank under accession numbers CP002921 (chromosome I), CP002922 (chromosome II), and CP002923 (pHH400).

The work was supported by grants from the National 863 Program of China (2009AA09Z401), the National Natural Science Foundation of China (30621005, 30830004, and 30925001), and Chinese Academy of Sciences (KSCX2-EW-G-2-4).

#### REFERENCES

1. Altschul, S. F., et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.
2. Cline, S. W., and W. F. Doolittle. 1992. Transformation of members of the genus *Haloarcula* with shuttle vectors based on *Halobacterium halobium* and *Haloferax volcanii* plasmid replicons. *J. Bacteriol.* **174**:1076–1080.
3. Cline, S. W., W. L. Lam, R. L. Charlebois, L. C. Schalkwyk, and W. F. Doolittle. 1989. Transformation methods for halophilic archaeobacteria. *Can. J. Microbiol.* **35**:148–152.
4. Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**:4636–4641.
5. Guo, F. B., H. Y. Ou, and C. T. Zhang. 2003. ZCURVE: a new system for recognizing protein-coding genes in bacterial and archaeal genomes. *Nucleic Acids Res.* **31**:1780–1789.
6. Han, J., Q. Lu, L. Zhou, H. Liu, and H. Xiang. 2009. Identification of the polyhydroxyalkanoate (PHA)-specific acetoacetyl coenzyme A reductase among multiple FabG paralogs in *Haloarcula hispanica* and reconstruction of the PHA biosynthetic pathway in *Haloferax volcanii*. *Appl. Environ. Microbiol.* **75**:6168–6175.
7. Han, J., Q. Lu, L. Zhou, J. Zhou, and H. Xiang. 2007. Molecular characterization of the *phaEC<sub>Hm</sub>* genes, required for biosynthesis of poly(3-hydroxybutyrate) in the extremely halophilic Archaeon *Haloarcula marismortui*. *Appl. Environ. Microbiol.* **73**:6058–6065.
8. Juez, G., F. Rodriguezvalera, A. Ventosa, and D. J. Kushner. 1986. *Haloarcula hispanica* spec. nov. and *Haloferax gibbonsii* spec. nov., two new species of extremely halophilic archaeobacteria. *Syst. Appl. Microb.* **8**:75–79.
9. Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**:D277–D280.
10. Lagesen, K., et al. 2007. RNAmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**:3100–3108.
11. Liu, H., J. Han, X. Liu, J. Zhou, and H. Xiang. 2011. Development of *pyrF*-based gene knockout systems for genome-wide manipulation of the archaea *Haloferax mediterranei* and *Haloarcula hispanica*. *J. Genet. Genomics* **38**:261–269.
12. Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
13. Porter, K., B. E. Russ, and M. L. Dyll-Smith. 2007. Virus-host interactions in salt lakes. *Curr. Opin. Microbiol.* **10**:418–424.
14. Tettelin, H., D. Radune, S. Kasif, H. Khouri, and S. L. Salzberg. 1999. Optimized multiplex PCR: efficiently closing a whole-genome shotgun sequencing project. *Genomics* **62**:500–507.