JB Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Sequence of *Halorubrum* sp. Strain T3, an Extremely Halophilic Archaeon Harboring a Virus-Like Element

Shaoxing Chen,[a,b,c] Chuanming Wang,[c,d] Zhiwei Zhao,[b] and Zhu L. Yang[a]

Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, People's Republic of China[a]; Key Laboratory of Conservation and Utilization for Bioresources and Key Laboratory of Microbial Diversity in Southwest China, Ministry of Education, Yunnan University, Kunming, People's Republic of China[b]; University Key Laboratory of Crop High Quality and High Effective Cultivation and Safety Control in Yunnan Province, Honghe University, Mengzi, People's Republic of China[c]; and Modern Biological Research Center, Yunnan University, Kunming, People's Republic of China[d]

***Halorubrum* sp. strain T3, harboring a virus-like element, was isolated from a sample collected from a solar saltern in Yunnan, China. Several strains of *Halorubrum* pleomorphic viruses were reported in this genus recently; however, the virus-host interaction in haloarchaea remains unclear. To explore this issue, here we present the genome sequence of *Halorubrum* sp. strain T3 (3,168,011 bp, 68.48% G+C content).**

Species of the genus *Halorubrum* are widely distributed, mainly in a hypersaline environment (14). Recently, four strains of *Halorubrum* pleomorphic viruses (HRPV-1, HRPV-2, HRPV-3 and HRPV-6) were reported (17, 18). However, the virus-host interaction in haloarchaea remains unclear. In our laboratory, tens of strains isolated from the Mohei solar saltern (Yunnan, People's Republic of China) have been shown to belong to the genus *Halorubrum*; however, only a single strain, *viz*, strain T3, contains a virus-like element, which is largely homologous with HRPV-3 (our unpublished data). The 16S rRNA sequence of strain T3 has a close relationship to *Halorubrum chaoviator* Halo-G^T (99.217% similarity) (15) and *Halorubrum trapanicum* NRC 34021^T (99.215%) (20). It can grow well in universal modified growth medium (MGM) under 37°C and can be easily transformed in laboratory without polyethylene glycol (PEG) mediation (2). Compared with *Halorubrum lacusprofundi* ATCC 49239, from which virus was absent, which has genome sequence accession no. CP001365.1 to CP001367.1 and was isolated from a deep lake in Antarctica, strain T3 was isolated from a typical hypersaline surroundings. To explore the virus-host interaction, we sequenced the genome of *Halorubrum* sp. strain T3.

The genome sequence of *Halorubrum* sp. strain T3 was determined by a whole-genome shotgun strategy using Illumina Hiseq2000 paired-end sequencing technology (16), a 500-bp-span paired-end library (~300 Mb of available reads with ~100-fold genome coverage), and a 2,000-bp-span paired-end library (~150 Mb of available reads with ~50-fold genome coverage). All reads were assembled into 16 contigs (>919 bp in size) and 10 scaffolds (>2,381 bp in size) by use of SOAP *de novo* v1.05 software (11). Genes were predicted by using Glimmer 3.0 (4). Noncoding RNAs (rRNA and tRNA) were determined by using tRNAscan-SE (13), RNAmmer (10), and the Rfam database (6). The repeat sequence annotation was obtained by using the RepeatMasker software and Repbase database (9). The gene annotation was achieved by BLAST (1), the Swiss-Prot and Clusters of Orthologous Groups (COG) databases (19), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (8). The prephage was predicted by using Prohinder software and the ACLAME database (12). Finally, clustered regularly interspaced short palindromic repeats (CRISPRs) were predicted by using CRISPRFinder software (7).

Halorubrum sp. strain T3 has a genome size of 3,168,011 bp (scaffold length) and 68.48% G+C content. A total of 3,369 coding sequences (CDSs) and 48 tRNAs were identified. Two rRNA operons, no small RNAs (sRNA), no potential prephages, and 2,630 tandem repeats were found. Among the 3,369 genes identified, 1,577 were classified into 21 certain functional COG sets or 1,563 were classified into 30 certain functional KEGG sets. Eighteen possible CRISPRs, resembling the mammalian adaptive immune system (3, 5), were found on the genome sequence. Analysis of the CRISPR-rich genome sequence of strain T3 and comparison with other genomes will provide further insights into the genetics and virus-host interaction of haloarchaea.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited in the DDBJ, EMBL, and GenBank databases under accession no. ALWQ00000000. The version described in this article is the first version, ALWQ01000000. The accession number for the 16S rRNA gene is JQ936845.

ACKNOWLEDGMENTS

This work was supported by the Joint Funds of the National Natural Science Foundation of China and Yunnan Provincial Government (no. U0836604) and the CAS/SAFEA International Partnership Program for Creative Research Teams.

REFERENCES

1. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
2. **Chen S, et al.** 2012. Lateral gene transfer occurring in haloarchaea: an interpretative imitation study. World J. Microbiol. Biotechnol. **28**:2913–2918.
3. **Datsenko KA, et al.** 2012. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat. Commun. **3**:945. doi:10.1038/ncomms1937.
4. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.

Received 22 August 2012 **Accepted** 17 September 2012

Address correspondence to Zhu L. Yang, fungi@mail.kib.ac.cn.

doi:10.1128/JB.01528-12

5. **Goren M, Yosef I, Edgar R, Qimron U.** 1 May 2012. The bacterial CRISPR/Cas system as analog of the mammalian adaptive immune system. RNA Biol. [Epub ahead of print.] doi:org/10.4161/rna.20177.
6. **Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR.** 2003. Rfam: an RNA family database. Nucleic Acids Res. **31**:439–441.
7. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35**:52–57.
8. **Kanehisa M, Goto S.** 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. **28**:27–30.
9. **Kohany O, Gentles A, Hankus L, Jurka J.** 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics **7**:474. doi:10.1186/1471-2105-7-474.
10. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
11. **Li R, Li Y, Kristiansen K, Wang J.** 2008. SOAP: Short Oligonucleotide Alignment Program. Bioinformatics **24**:713–714.
12. **Lima-Mendez G, Van Helden J, Toussaint A, Leplae R.** 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. Bioinformatics **24**:863–865.
13. **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.
14. **Luque R, González-Domenech CM, Llamas I, Quesada E, Béjar V.** 2012. Diversity of culturable halophilic archaea isolated from Rambla Salada, Murcia (Spain). Extremophiles **16**:205–213.
15. **Mancinelli RL, et al.** 2009. *Halorubrum chaoviator* sp. nov., a haloarchaeon isolated from sea salt in Baja California, Mexico, Western Australia and Naxos, Greece. Int. J. Syst. Evol. Microbiol. **59**:1908–1913.
16. **Matochko W, et al.** 20 July 2012. Deep sequencing analysis of phage libraries using Illumina platform. Methods [Epub ahead of print.] http://dx.doi.org/10.1016/j.ymeth.2012.07.006.
17. **Pietilä MK, Roine E, Paulin L, Kalkkinen N, Bamford DH.** 2009. An ssDNA virus infecting archaea: a new lineage of viruses with a membrane envelope. Mol. Microbiol. **72**:307–319.
18. **Sencilo A, Paulin L, Kellner S, Helm M, Roine E.** 2012. Related haloarchaeal pleomorphic viruses contain different genome types. Nucleic Acids Res. **40**:5523–5534.
19. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. **28**:33–36.
20. **Ventosa A, Gutiérrez MC, Kamekura M, Dyall-Smith ML.** 1999. Proposal to transfer *Halococcus turkmenicus*, *Halobacterium trapanicum* JCM 9743 and strain GSL-11 to *Haloterrigena turkmenica* gen. nov., comb. nov. Int. J. Syst. Evol. Microbiol. **49**:131–136.