Complete Genome Sequence of a Halophilic Bacterium, *Halomonas* sp. Strain NyZ770, from Mariana Trench Sediment

Xinting Lai,[a] Ying Xu[a]

aState Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic & Developmental Sciences and School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, Shanghai, China

ABSTRACT *Halomonas* sp. strain NyZ770 is a bacterium that was isolated from Mariana Trench sediment. Here, the complete genome sequence of this strain is reported. The genome was sequenced with the Illumina NovaSeq and Pacific Biosciences Sequel sequencing platforms and consists of a single chromosome of 4,024,853 bp, with a G+C content of 60.21%.

*H*alomonas species are moderately halophilic bacteria that have been reported to be distributed in a variety of saline environments, even deep-sea environments (1, 2). Here, we report the complete genome sequence of *Halomonas* sp. strain NyZ770, a bacterium from the hadal trench, to better understand deep-sea microbiology. *Halomonas* sp. strain NyZ770 was isolated from hadal trench sediment that had been collected at 6,300 m, at a depth of 16 to 20 cm below the seafloor, from the Challenger Deep in the Mariana Trench (10.8895°N, 142.2260°E). The sediment sample was enriched in artificial seawater (ASW) basal medium, a modified ASW medium (3) omitting carbon sources, with 30 mM glucose and 10 mM *N*-acetylglucosamine under *in situ* pressure of 60 MPa for 6 months and then was selectively enriched in ASW basal medium with 1 mM 4-hydroxybenzoate (4HBA) for an additional 4 months. Finally, serial dilution on ASW agar plates with 4HBA under ambient pressure (0.1 MPa) yielded an isolate designated strain NyZ770. The 16S rRNA gene from strain NyZ770 exhibited the greatest identity (>99%) to those from three strains of the genus *Halomonas* (GenBank accession numbers AB362301.1, MN258543.1, and CP023656.1), indicating that it should be classified as a *Halomonas* strain.

For genomic sequencing, a single colony of strain NyZ770 was cultured overnight in liquid lysogeny broth (LB) (4, 5) before the cells were washed twice and harvested. The genomic DNA was extracted by using the cetyltrimethyl ammonium bromide (CTAB) (6) method with minor modifications. The NovaSeq library was prepared using the TruSeq DNA sample preparation kit (Illumina, USA) with an insert size of 400 bp, in paired-end (2 × 150-bp) sequencing mode. The Pacific Biosciences (PacBio) sequencing library was prepared using the template preparation kit v1.0 (PacBio, USA) with an insert size of 20 kb, in standard sequencing mode. Genome sequencing was performed by Personal Biotechnology Co. (Shanghai, China) using the Illumina NovaSeq and PacBio Sequel platforms.

For the Illumina data, the number of reads is 8,702,654 reads in total. For the PacBio data, the read N_{50} is 11,406 bp, with a total sequence number of 245,204 sequences. AdapterRemoval (v2.2.2) (7) and SOAPec (v2.03) (8) were used to remove the joint contamination and to filter the error reads, respectively. The resulting data from the PacBio sequencing platform were assembled with HGAP (v4) (9) and Canu (v1.7.1) (10) software to obtain the contig sequence. The Illumina data were used by Pilon (v1.18) (11) to correct the resulting PacBio assembly. For the circularization step, the overlap was identified and trimmed with Circlator (v1.5.5) (12). GeneMarkS (v4.32) (13) was

used for gene prediction for the complete genome. tRNAscan-SE (v1.3.1) (14) was used to predict tRNAs, and Barrnap software (v0.9) (15) was used to predict rRNAs (5S rRNAs, 16S rRNAs, and 23S rRNAs). Default parameters were used for all software unless otherwise noted. The protein-coding genes were annotated with the NCBI Non-Redundant Protein Sequence (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG), eggNOG, and Swiss-Prot databases.

The complete genome of this strain consisted of a single chromosome of 4,024,853 bp, with a G+C content of 60.21%. The number of open reading frames (ORFs) was 3,690, the total ORF length was 3,515,844 bp, the average ORF length was 952.80 bp, the ORF/genome value (coding percentage) was 87.35%, and the G+C content in the ORF region was 61.12%. Genome annotation suggested that strain NyZ770 has 3,542, 2,174, and 2,790 protein-coding genes according to the NR, KEGG, and Swiss-Prot databases, respectively, which accounted for 95.99%, 58.92%, and 75.61% of the total number of predicted ORFs, respectively. Regarding noncoding RNA prediction, the tRNA, 5S rRNA, 16S rRNA, and 23S rRNA copy numbers were 60, 6, 6, and 6, respectively.

This assembly will facilitate genome-wide comparison studies with a focus on the ecology of bacteria living in hadal trench environments under high hydrostatic pressures.

Data availability. The genome sequence and annotation data for *Halomonas* sp. strain NyZ770 were deposited in GenBank under BioProject number PRJNA769991, BioSample number SAMN22187422, and GenBank accession number CP085143. The raw Illumina and PacBio data for strain NyZ770 were deposited in the Sequence Read Archive (SRA) under BioProject number PRJNA769991 and SRA accession number SRS11013316.

REFERENCES

1. Kaye J, Baross J. 2000. High incidence of halotolerant bacteria in Pacific hydrothermal-vent and pelagic environments. FEMS Microbiol Ecol 32:249–260. https://doi.org/10.1111/j.1574-6941.2000.tb00718.x.
2. Tsurumaki M, Deno S, Galipon J, Arakawa K. 2019. Complete genome sequence of halophilic deep-sea bacterium *Halomonas axialensis* strain Althf1. Microbiol Resour Announc 8:e00839-19. https://doi.org/10.1128/MRA.00839-19.
3. Yang S, Lv Y, Liu X, Wang Y, Fan Q, Yang Z, Boon N, Wang F, Xiao X, Zhang Y. 2020. Genomic and enzymatic evidence of acetogenesis by anaerobic methanotrophic archaea. Nat Commun 11:3941. https://doi.org/10.1038/s41467-020-17860-8.
4. Bertani G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J Bacteriol 62:293–300. https://doi.org/10.1128/jb.62.3.293-300.1951.
5. Bertani G. 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol 186:595–600. https://doi.org/10.1128/JB.186.3.595-600.2004.
6. Clarke JD. 2009. Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. Cold Spring Harb Protoc 2009:pdb.prot5177. https://doi.org/10.1101/pdb.prot5177.
7. Schubert M, Lindgreen S, Orlando L. 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 9:88. https://doi.org/10.1186/s13104-016-1900-2.
8. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18.
9. Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods 13:1050–1054. https://doi.org/10.1038/nmeth.4035.
10. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
11. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
12. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.
13. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes: implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618. https://doi.org/10.1093/nar/29.12.2607.
14. 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.
15. Liu N, Zhu L, Zhang Z, Huang H, Jiang L. 2019. Draft genome sequence of a multidrug-resistant *bla*$_{OXA-69}$-producing *Acinetobacter baumannii* L13 isolated from Tarim River sample in China. J Glob Antimicrob Resist 18:145–147. https://doi.org/10.1016/j.jgar.2019.05.022.