Complete Genome Sequence of "*Candidatus* Hydrogeosomobacter endosymbioticus," an Intracellular Bacterial Symbiont of the Anaerobic Ciliate Scuticociliate GW7

Yasuo Shiohama,[a] Kazutaka Takeshita,[b] Yuga Hirakata,[c] Masaru K. Nobu,[c] Michihiro Ito,[a] ⓘ Naoya Shinzato[a,c]

[a]Tropical Biosphere Research Center, University of the Ryukyus, Nishihara, Okinawa, Japan
[b]Faculty of Bioresource Sciences, Akita Prefectural University, Akita, Japan
[c]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

ABSTRACT The bacterium "*Candidatus* Hydrogenosomobacter endosymbioticus" is an intracellular symbiont of anaerobic scuticociliate GW7, which is associated with hydrogenosome together with methanogenic archaea. Here, we report a complete genome sequence of the symbiont consisting of 827 kbp. Knowing this sequence would contribute to the understanding of the metabolic interactions and evolution of the tripartite symbiosis.

The free-living anaerobic ciliate scuticociliate GW7 isolated from sewage treatment tank harbors a symbiotic bacterium named "*Candidatus* Hydrogenosomobacter endosymbioticus," a *Holospora*-related alphaproteobacterium (1). The symbiotic bacterium was closely associated with a hydrogen-producing organelle, hydrogenosome, together with *Methanoregula*-related methanogenic archaeal symbiont (2, 3). In this announcement, we attempted to determine the whole genome sequence of "*Ca*. Hydrogenosomobacter endosymbioticus" by metagenomic approach.

The cells of "*Ca*. Hydrogenosomobacter endosymbioticus" were fractionated from a total of 4 L of ciliate cultures as described previously (1, 4–6). After starvation for 18 h in fresh medium, the ciliate cells were homogenized with a tissue grinder. The symbiont cells were fractionated by passing through a 5-μm membrane filter (Millipore) to remove ciliate cell debris and subjected to DNA extraction using REDExtract-N-Amp tissue PCR kit (Sigma-Aldrich). Finally, 25 ng of DNA was used for transposome-mediated library preparation using Nextera DNA Flex library prep kit (Illumina). The library was sequenced using MiSeq sequencer (Illumina) with MiSeq reagent kit v3 (Illumina).

The obtained reads were assembled as described below with default parameters except where otherwise noted. The obtained reads were trimmed by 33 of phred score and the reads less than 35 bp were discarded using Trimmomatic v0.36 (7). A total of 28,874,984 reads were assembled using SPAdes v3.13.0 (8). The obtained contigs were clustered by binning using CONCOCT v1.1 (9), Maxbin2 v2.2.4 (10), and MetaBAT2 v1.7 (11). Finally, 22 consensus bins were classified using GTDB-Tk classify v1.1.0 (12). Based on the results, we determined that one of the bins was the genome of "*Ca*. Hydrogenosomobacter endosymbioticus." This genome bin contained 88 contigs, and an N_{50} was 825,562 bp. To examine the circularity of the longest contig (825,562 bp), the symbiont genome was amplified by PCR with a set of primers, 5′-CCTGCTCTGGCGTTAATTGTTGTG-3′ and 5′-CCGCTCTTCTTC CGTCAGCCTTCCATAG-3′, which were designed for each end of the longest contig. The PCR product (15,367 bp) was inserted into pCR4-TOPO vector (Thermo Fisher Scientific) and sequenced by primer walking. The obtained sequences were aligned to the longest contig, and finally, a circular chromosome of 826,669 bp was obtained. The genome completeness estimated by CheckM was 89.6%, which is a program for assessing the genome completeness based on the retention rate of marker genes (13). The sequencing depth

estimated by mapping reads to circular genome was 81×. The open reading frame (ORF) detection and gene annotation have been performed using DFAST pipeline v1.2.13 (14, 15). As the results, 721 protein-coding sequences, 45 tRNAs, and 3 rRNAs were detected. The GC content was 41.3%.

The reported genome sequence will make clear the physiological features of "*Ca.* Hydrogenosomobacter endosymbioticus" and help deepen understanding of the microbial symbiotic association in anaerobic environments.

Data availability. The genome sequence of "*Ca.* Hydrogenosomobacter endosymbioticus" has been deposited in the DNA Data Bank of Japan (DDBJ) under accession number AP025225. The associated BioProject, BioSample, and SRA accession numbers are PRJDB12495, SAMD00412742, and DRA013104, respectively.

ACKNOWLEDGMENT

REFERENCES

1. Takeshita K, Yamada T, Kawahara Y, Narihiro T, Ito M, Kamagata Y, Shinzato N. 2019. Tripartite symbiosis of an anaerobic scuticociliate with two hydrogenosome-associated endosymbionts, a *Holospora*-related alphaproteobacterium and a methanogenic archaeon. Appl Environ Microbiol 85:e00854-19. https://doi.org/10.1128/AEM.00854-19.
2. Fenchel T, Finlay BJ. 2010. Free-living protozoa with endosymbiotic methanogens, p 1–11. *In* Hackstein J (ed), (Endo)symbiotic methanogenic archaea. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-13615-3_1.
3. Shinzato N, Takeshita K, Kamagata Y. 2018. Methanogenic and bacterial endosymbionts of free-living anaerobic ciliates, p 37–53. *In* Hackstein J (ed), (Endo)symbiotic methanogenic archaea. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-319-98836-8_4.
4. Yamada K, Kamagata Y, Nakamura K, Inamori Y, Nakamura I. 1994. Selectivity of food bacteria for the growth of anaerobic ciliate *Trimyema compressum*. Arch Microbiol 161:229–233. https://doi.org/10.1007/BF00248697.
5. Shinzato N, Watanabe I, Meng XY, Sekiguchi Y, Tamaki H, Matsui T, Kamagata Y. 2007. Phylogenetic analysis and fluorescence in situ hybridization detection of archaeal and bacterial endosymbionts in the anaerobic ciliate *Trimyema compressum*. Microb Ecol 54:627–636. https://doi.org/10.1007/s00248-007-9218-1.
6. Shinzato N, Aoyama H, Saitoh S, Nikoh N, Nakano K, Shimoji M, Shinzato M, Satou K, Teruya K, Hirano T, Yamada T, Nobu MK, Tamaki H, Shirai Y, Park S, Narihiro T, Liu WT, Kamagata Y. 2016. Complete genome sequence of the intracellular bacterial symbiont TC1 in the anaerobic ciliate *Trimyema compressum*. Genome Announc 4:e01032-16. https://doi.org/10.1128/genomeA.01032-16.
7. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
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. https://doi.org/10.1089/cmb.2012.0021.
9. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103.
10. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.
11. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. Peer J 7:e7359. https://doi.org/10.7717/peerj.7359.
12. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
13. 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/10.1101/gr.186072.114.
14. Tanizawa Y, Fujisawa T, Kaminuma E, Nakamura Y, Arita M. 2016. DFAST and DAGA: web-based integrated genome annotation tools and resources. Biosci Microbiota Food Health 35:173–184. https://doi.org/10.12938/bmfh.16-003.
15. Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34:1037–1039. https://doi.org/10.1093/bioinformatics/btx713.