Complete Genome Sequence of *Halomonas* sp. Strain SH5A2, a Dye-Degrading Halotolerant Bacterium Isolated from the Salinas and Aguada Blanca National Reserve in Peru

Walter F. Manya,[a] Wendy C. Lizárraga,[a] Carlo G. Mormontoy,[a] Mario A. Taira,[a] Pablo S. Ramírez[a]

[a]Laboratory of Molecular Microbiology and Biotechnology, Faculty of Biological Sciences, Universidad Nacional Mayor de San Marcos, Lima, Peru

ABSTRACT *Halomonas* sp. strain SH5A2 is a halotolerant bacterium isolated from Salinas Lake at 4,300 m above sea level in Peru. Here, we report its complete genome sequence with a length of 3,849,224 bp and highlight the presence of genes putatively related to dye degradation, such as NADPH-dependent oxidoreductases.

*H*alomonas strains are Gram-negative, rod-shaped, aerobic, and halotolerant bacteria (1). Many strains grow and decolorize harmful synthetic azo dyes in anaerobic cultures with up to 20% (wt/vol) NaCl (2–4). Moreover, enzymes with azoreductase function were characterized in this genus (5, 6).

A novel strain of *Halomonas* (SH5A2) was isolated in plates with seawater medium (pH 7.5) with 12.5% NaCl from a 10% precultured sample in seawater broth (7) collected from Salinas Lake (16°20′10″S, 71°08′30″W) within the Salinas and Aguada Blanca National Reserve in Arequipa (Peru) and incubated at 30°C under static conditions for 48 h.

For DNA extraction, strain SH5A2 was cultivated in seawater broth at 30°C under static conditions for 48 h to the mid-log phase (8). This culture was used for DNA extraction with the Wizard genomic DNA purification kit (Promega, USA) following the manufacturer's instructions. Library construction prepared inserts of 20 kb for single-molecule real-time (SMRT) sequencing on the Pacific Bioscience RS II platform. Raw sequencing data included 116,274 subreads encompassing 1,272,266,623 bp. The average subread length and N_{50} value were 10,941 bp and 16,510 bp, respectively. The Canu v2.0 assembler (9) corrected and trimmed reads to produce 5,379 reads with 129,094,014 bp, which were used to perform *de novo* assembly into a single contig (correctedErrorRate=0.039). Subsequently, the Arrow algorithm in the GenomicConsensus v2.3.3 package (Pacific Biosciences, USA) was chosen to improve genomic sequence quality through several polishing rounds. The genome circularization was performed with Circlator v1.5.5 using corrected reads and the polished assembly (10) and visualized with the CGView Server (11).

TABLE 1 Summary statistics and characteristic features of strain SH5A2 genome sequencing

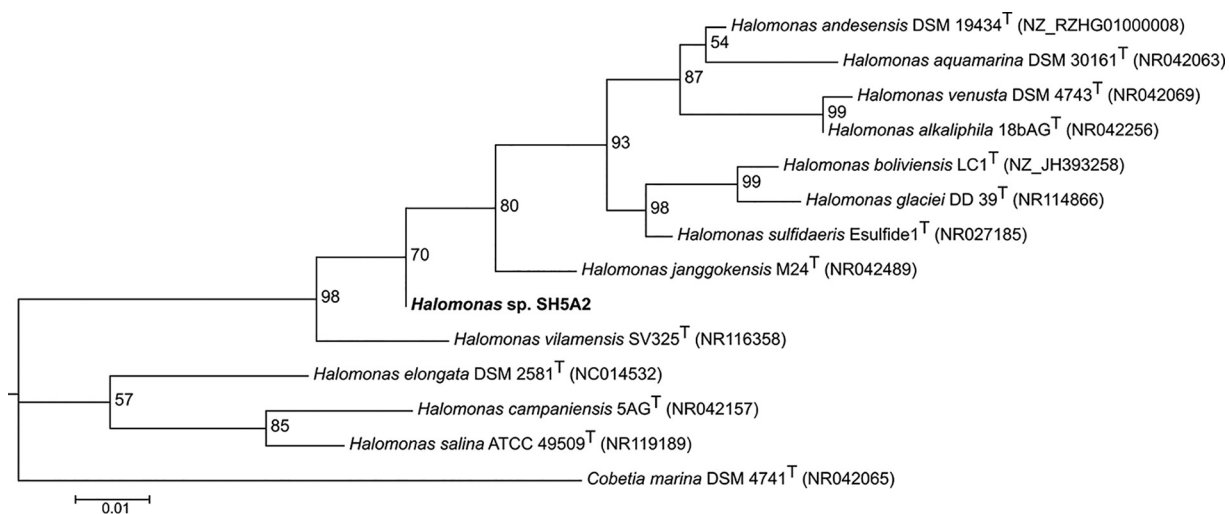| Characteristic	Data for *Halomonas* sp. SH5A2
BioProject accession no.	PRJNA639301
BioSample accession no.	SAMN15395493
No. of subreads	116,274
Subread N_{50} (bp)	16,510
Genome size (bp)	3,849,224
G+C content (%)	56.9
Coverage (×)	33.44
Total no. of genes	3,578
No. of coding sequences	3,495
No. of rRNAs	18
No. of tRNAs	64

FIG 1 Phylogenetic tree of the 16S rRNA gene in *Halomonas* sp. SH5A2 based on the maximum likelihood method with 1,000 bootstraps. GenBank accession numbers are listed in parentheses.

Annotation was performed with the RAST v2.0 server (12), Prokka v1.14.0 (13), and NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v4.12 (14). In addition, to perform a phylogenetic analysis, a nucleotide sequence alignment was made with MAFFT v7.471 (https://mafft.cbrc.jp/alignment/software), followed by the best nucleotide substitution model prediction (HKY+F+R2) and phylogenetic tree construction using the 16S rRNA gene with IQ-TREE v1.6.12 software (15). Finally, tree visualization was performed with FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree). Default parameters were used except where otherwise noted.

Summary statistics and characteristic features of the complete genome sequence of *Halomonas* sp. strain SH5A2 are given in Table 1. This genome mainly encodes proteins related to carbohydrate metabolism, and it has 46 proteins related to xenobiotic metabolism and biodegradation. Genes encoding proteins putatively related to dye degradation were predicted as NADPH-dependent oxidoreductases (HXW73_11090 and HXW73_16825) and a pseudogene for NAD(P)H-dependent oxidoreductase protein (HXW73_07605) previously identified by RAST as a flavin mononucleotide (FMN)-dependent NADH-azoreductase. The phylogenetic tree shown in Fig. 1 allows us to determine that strain SH5A2 diverges significantly from other species, forming a separate group that could belong to a new species. We conclude that *Halomonas* sp. SH5A2 could have great potential in bioremediation of textile effluents due to its capacity to tolerate high NaCl concentrations and the presence of genes putatively involved in dye degradation.

Data availability. The genome sequence and raw data reported here were deposited in the GenBank (accession number CP058321) and SRA (accession number SRR12147485) databases, respectively.

REFERENCES

1. Dobson SJ, Franzmann PD. 1996. Unification of the genera *Deleya* (Baumann et al. 1983), *Halomonas* (Vreeland et al. 1980), and *Halovibrio* (Fendrich 1988) and the species *Paracoccus halodenitrificans* (Robinson and Gibbons 1952) into a single genus, *Halomonas*, and placement of the genus *Zymobacter* in the family Halomonadaceae. Int J Syst Bacteriol 46:550–558. https://doi.org/10.1099/00207713-46-2-550.

2. Asad S, Amoozegar MA, Pourbabaee AA, Sarbolouki MN, Dastgheib SMM. 2007. Decolorization of textile azo dyes by newly isolated halophilic and halotolerant bacteria. Bioresour Technol 98:2082–2088. https://doi.org/10.1016/j.biortech.2006.08.020.

3. Guo J, Zhou J, Wang D, Tian C, Wang P, Uddin MS. 2008. A novel moderately halophilic bacterium for decolorizing azo dye under high salt condition. Biodegradation 19:15–19. https://doi.org/10.1007/s10532-007-9110-1.

4. Guadie A, Gessesse A, Xia S. 2018. *Halomonas* sp. strain A55, a novel dye decolorizing bacterium from dye-uncontaminated Rift Valley Soda Lake. Chemosphere 206:59–69. https://doi.org/10.1016/j.chemosphere.2018.04.134.

5. Eslami M, Amoozegar MA, Asad S. 2016. Isolation, cloning and characterization of an azoreductase from the halophilic bacterium *Halomonas elongata*. Int J Biol Macromol 85:111–116. https://doi.org/10.1016/j.ijbiomac.2015.12.065.

6. Tian F, Guo G, Zhang C, Yang F, Hu Z, Liu C, Wang SW. 2019. Isolation, cloning and characterization of an azoreductase and the effect of salinity on its expression in a halophilic bacterium. Int J Biol Macromol 123:1062–1069. https://doi.org/10.1016/j.ijbiomac.2018.11.175.

7. Mullakhanbhai MF, Larsen H. 1975. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. Arch Microbiol 104:207–214. https://doi.org/10.1007/BF00447326.

8. Dyall-Smith M. 2009. The Halohandbook: protocols for haloarchaeal genetics. University of Melbourne, Melbourne, Australia.

9. 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.

10. Hunt M, De Silva N, 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.

11. Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36:W181–W184. https://doi.org/10.1093/nar/gkn179.

12. Aziz RK, Bartels D, Best A, 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. https://doi.org/10.1186/1471-2164-9-75.

13. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

14. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

15. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.