Metagenome-Assembled Genome of *Halomonas* sp. Isolate SL48-SHIP-3 from the Microbial Mat of Salt Lake Number 48 (Novosibirsk Region, Russia)

Aleksandra A. Shipova,[a] Anton V. Korzhuk,[a] Valeria Shlyahtun,[a] Alla V. Bryanskaya,[a] Aleksey S. Rozanov,[a] Sergey E. Peltek[a]

[a]Federal Research Center–Institute of Cytology and Genetics of the Siberian Branch of the RAS, Novosibirsk, Russia

ABSTRACT The *Halomonas* sp. isolate SL48-SHIP-3 genome was obtained from metagenomics sequencing of the microbial mat of Salt Lake Number 48 (54.201806N, 78.179194E; Novosibirsk region, Russia). The sequenced and annotated genome is 2,575,909 bp and encodes 2,368 genes.

*H*alomonas is one of the genera of the family *Halomonadaceae* (*Proteobacteria*). These bacteria are halophiles that were previously discovered in other regions of the planet, including salt lakes (1).

Salt Lake Number 48 (54.201806N, 78.179194E; Novosibirsk Region, Russia) is a chloride sulfate lake (pH 8.0), with salinity ranging from 190 g/liter to 230 g/liter depending on weather conditions (2). A sample of the microbial mat, which consists of floating particles 2 to 10 mm in size, was taken from the coastal part of this lake and stored in alcohol at −70°C. From this sample, a metagenome was obtained, from which the genomes of individual microorganisms were then isolated. We have already described one of the genomes in a previous article (3). Total DNA was isolated using the NucleoSpin soil kit (genomic DNA from soil) with the default protocol. Libraries for metagenome sequencing were prepared at the Center of Genomic Studies, Institute of Cytology and Genetics of the Siberian Branch of the RAS (ICG SB RAS), using a NEBNext Ultra DNA library prep kit for Illumina; the insert size was 450 bp. Metagenome paired-end sequencing was performed on a NovaSeq platform (Illumina) at Genetico LLC using the NovaSeq 6000 S2 reagent kit (200 cycles).

A total of 398,852,702 reads were sequenced; the average read length was 100 bp. The reads were processed by Trimmomatic version 0.36 (using options MINLEN:95 and CROP:97) (4). *De novo* assembly of short reads into scaffolds was performed using SPAdes version 3.11.1 with the option "-only-assembler" (5). Contigs shorter than 1,000 bp were deleted. Binning of metagenomics scaffolds into separate clusters (in which one bin represents one genome) was carried out using MaxBin (version 2.2.4) with default parameters (6).

Here, we describe one of the resulting genomes with coverage of 629.5×. The classification of this genome to the family level was obtained using MetaWRAP version 1.2.2 (7) with the option "classify_bins." The genome belongs to the family *Halomonadaceae* (within the class *Gammaproteobacteria*). We aligned amino acid sequences (which were found by MaxBin for this cluster) against the NCBI nr protein database. Most of them had the best match (with the highest score and an E value of <0.05) with *Halomonas* sp. strain es.049. We compared this genome with the reference genomes of the most closely related members of the genus *Halomonas* (that had a match with the described genome) using the Average Nucleotide Identity (ANI) calculator (http://enve-omics.ce.gatech.edu/ani/) with default parameters. The genomes of *Halomonas arcis* strain CGMCC (identity, 77.78%), *Halomonas campaniensis* strain LS21 (identity, 78.04%), and *Halomonas* sp. es.049 (identity, 80.18%) were closest to our genome. These results

suggest that our organism may constitute a new strain. The genome was checked for contamination using CheckM version 1.0.13 (8) (with the option "taxonomy_wf family Halomonadaceae"); the contamination rate was 0.64%, and the completeness rate was 92.18%.

The genome is 2,575,909 bp long, consists of 149 contigs, and has a GC content of 56.79% and an N_{50} value of 37,322 bp (evaluated using QUAST version 5.0.2 with default parameters [9]). Open reading frame (ORF) prediction and automatic annotation were performed using NCBI PGAP (version 4.8) with the default parameters (10). The complete genome sequence contains 2,446 genes, 2,388 coding sequences (CDS), 1 rRNA (23S), 53 tRNAs, and 4 noncoding RNAs (ncRNAs).

Data availability. The raw metagenomics data have been deposited at DDBJ/EMBL/GenBank under accession no. SRR7943696. The draft genome sequence for *Halomonas* sp. isolate SL48-SHIP-3 has been deposited at DDBJ/EMBL/GenBank under accession no. VMQS00000000. The 149 contigs have been deposited under accession no. VMQS01000001 to VMQS01000149 (https://www.ncbi.nlm.nih.gov/Traces/wgs/VMQS01?display=contigs).

ACKNOWLEDGMENT

REFERENCES

1. Vreeland RH. 2015. Halomonas, p 1–19. *In* Bergey's manual of systematics of archaea and bacteria. John Wiley & Sons, Ltd, Chichester, United Kingdom.
2. Bryanskaya AV, Malup TK, Lazareva EV, Taran OP, Rozanov AS, Efimov VM, Peltek SE. 2016. The role of environmental factors for the composition of microbial communities of saline lakes in the Novosibirsk Region (Russia). BMC Microbiol 16(Suppl 1):4. https://doi.org/10.1186/s12866-015-0618-y.
3. Rozanov AS, Shipova AA, Bryanskaya AV, Peltek SE. 2019. Metagenome-assembled genome sequence of *Phormidium* sp. strain SL48-SHIP, isolated from the microbial mat of Salt Lake Number 48 (Novosibirsk Region, Russia). Microbiol Resour Announc 8:e00651-19. https://doi.org/10.1128/MRA.00651-19.
4. 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.
5. 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.
6. 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.
7. Uritskiy GV, Diruggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158. https://doi.org/10.1186/s40168-018-0541-1.
8. 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.
9. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
10. 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.