



Metagenome-Assembled Genome Sequences Recovered from Epilithic River Biofilm in the Tama River, Japan

 Joval N. Martinez,^{a,b} Arisa Nishihara,^{a,c}  Shin Haruta^a

^aDepartment of Biological Sciences, Tokyo Metropolitan University, Hachioji, Tokyo, Japan

^bDepartment of Natural Sciences, University of St. La Salle, Bacolod City, Negros Occidental, Philippines

^cBioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

ABSTRACT Draft genome sequences of putatively novel bacteria were assembled from the metagenome of epilithic biofilm samples collected from the Tama River (Tokyo, Japan). The metagenome contains 44,630,724 sequences, 44,792 contigs, and 48% G+C content. Binning resulted in 31 metagenome-assembled genomes (MAGs) with $\geq 50\%$ completeness.

Epilithic biofilm sustains river ecosystems by producing organic substances and degrading organic matter. The main components of the riverbed biofilm in clear streams are microscopically identified as oxygenic photosynthetic organisms (1), but their metabolic potentials and other diverse coexisting microbes have not been studied well. Hirose et al. reported a PCR amplicon analysis of riverbed biofilm from the Tama River, a major river flowing through the Tokyo area, Japan, and found an unexpected diversity of anoxygenic phototrophic bacteria (2, 3). Here, we report the metagenome-assembled genome sequences (MAGs) retrieved from metagenomic reads of the biofilm in the Tama River.

Epilithic biofilms that had developed on a stone in the riverbed were collected from the Tama River, Ome City, Tokyo, Japan (35°47'10.6"N, 139°15'16.5"E), on 23 November 2014. Biofilms were scraped off the stone, placed into one 1.5-ml collection tube using a sterilized cotton swab, stored on ice during transportation to the laboratory (1 h), and stored at -20°C until further use. DNA from the biofilm sample was extracted and purified using the PowerBiofilm DNA isolation kit (Qiagen) for metagenomic sequencing. Sequencing libraries were prepared using the Illumina TruSeq library prep kit. A total of 90,614,554 bp metagenome reads from paired-end sequencing (2×101 cycles) were quality filtered using the Illumina chastity filter (Hokkaido System Science Co., Ltd.) and assessed using FastQC v.1.1.1 (4). Metagenome analyses were implemented using DOE Systems Biology Knowledgebase (KBase) (5). A total of 44,630,724 sequences were retrieved from the trimming process using Trimmomatic v.0.32 (6). These were assembled as contigs using metaSPAdes v.3.13.0 (7). Quality assessment using QUAST v.4.4 (8, 9) resulted in a total of 44,792 contigs with 48% G+C content. The contigs were binned and optimized using Concoct v.1.3.4 (10). Optimization of binning resulted in 71 putative draft metagenome-assembled genome sequences (MAGs), in which 31 bacterial MAGs were $\geq 50\%$ complete with mostly $\leq 3\%$ contamination as determined using CheckM (11). Default parameters were employed for all the software used.

Based on taxonomic assignments using the Genome Taxonomy Database Toolkit (GTDB-Tk) Classify v.0.1.4 (12), 31 river biofilm (RB) MAGs were classified as members of the phyla *Proteobacteria* (RB00, RB05, RB08, RB09, RB15, RB24, RB29, RB30, RB37, RB44, RB52, RB72, RB77, RB81, RB82, RB83, RB85, RB89, RB90), *Bacteroidota* (RB06, RB13, RB22, RB46, RB50, RB56, RB76, RB88), *Verrucomicrobiota* (RB28, RB53, RB61), and *Cyanobacteria* (RB71). All MAGs recovered had $\leq 97\%$ average nucleotide identity (ANI) compared to the closest cultured relatives, suggesting that the bacteria which harbored these MAGs are

Citation Martinez JN, Nishihara A, Haruta S. 2021. Metagenome-assembled genome sequences recovered from epilithic river biofilm in the Tama River, Japan. *Microbiol Resour Announc* 10:e00664-21. <https://doi.org/10.1128/MRA.00664-21>.

Editor Frank J. Stewart, Montana State University

Copyright © 2021 Martinez et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Joval N. Martinez, j.martinez@usls.edu.ph, or Shin Haruta, sharuta@tmu.ac.jp.

Received 1 July 2021

Accepted 23 August 2021

Published 23 September 2021

potentially novel species (13). The MAGs were annotated using Prokka v.1.14.5 (14). The annotated putative draft genome sequences contain *puflMC* genes encoding photosynthetic reaction centers, assigned to only two MAGs belonging to the classes *Alphaproteobacteria* (RB90) and *Gammaproteobacteria* (RB00). Genes encoding proteorhodopsin were also recovered from four MAGs representing *Bacteroidota/Chlorobi* (RB24, RB50, RB56, RB88), supporting findings for the widespread distribution of these light-driven proton pumps in freshwater ecosystems (15). As for the nitrogen metabolisms, no homologous gene involved in nitrogen fixation (*nifHDKEN*) was recovered in the metagenome. However, two denitrification-related genes, *narG* and *nosZ*, were retrieved from MAGs belonging to the classes *Gammaproteobacteria* (RB82) and *Bacteroidia* (RB56). Genes encoding reactive oxygen species (ROS) (e.g., superoxide dismutase, peroxidases) were recovered from 12 MAGs, aiding in protection and survival against oxidative stress, similar to other heterotrophic bacteria (16).

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the BioProject accession number [PRJNA668882](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA668882). The raw sequence reads are available in the Sequence Read Archive (SRA) under the accession numbers [SRR13089527](https://www.ncbi.nlm.nih.gov/sra/SRR13089527) to [SRR13089538](https://www.ncbi.nlm.nih.gov/sra/SRR13089538). The MAGs are accessible at <https://kbase.us/n/95585/4/>.

ACKNOWLEDGMENTS

This work was supported in part by funding from the Tokyu Foundation for Better Environment to S.H. J.N.M. was supported by the Photosynthetic Microbial Consortia Laboratory, donated by the Institute of Fermentation, Osaka (IFO).

We thank Setsuko Hirose for help with the sample collection.

REFERENCES

- Allan JD, Castillo MM. 2007. Stream ecology: structure and function of running waters, 2nd ed. Springer, Dordrecht, Netherlands. <https://doi.org/10.1007/978-1-4020-5583-6>.
- Hirose S, Nagashima KVP, Matsuura K, Haruta S. 2012. Diversity of purple phototrophic bacteria, inferred from *puflM* gene, within epilithic biofilm in Tama River, Japan. *Microbes Environ* 27:327–329. <https://doi.org/10.1264/jsme2.me11306>.
- Hirose S, Matsuura K, Haruta S. 2016. Phylogenetically diverse aerobic anoxygenic phototrophic bacteria isolated from epilithic biofilms in Tama River, Japan. *Microbes Environ* 31:299–306. <https://doi.org/10.1264/jsme2.ME15209>.
- Brown J, Pirrung M, McCue LA. 2017. FQC Dashboard: integrates FastQC results into a Web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics* 33:3137–3139. <https://doi.org/10.1093/bioinformatics/btx373>.
- Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, et al. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.
- 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>.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
- Mikheenko A, Valin G, Prijibelski A, Saveliev V, Gurevich A. 2016. Icarus: visualizer for de novo assembly evaluation. *Bioinformatics* 32:3321–3323. <https://doi.org/10.1093/bioinformatics/btw379>.
- 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>.
- 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>.
- 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>.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
- Kim M, Oh H-S, Park S-C, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* 64:346–351. <https://doi.org/10.1099/ijs.0.059774-0>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Atamna-Ismaeel N, Sabehi G, Sharon I, Witzel K-P, Labrenz M, Jürgens K, Barkay T, Stomp M, Huisman J, Beja O. 2008. Widespread distribution of proteorhodopsins in freshwater and brackish ecosystems. *ISME J* 2:656–662. <https://doi.org/10.1038/ismej.2008.27>.
- Diaz JM, Hansel CM, Voelker BM, Mendes CM, Andeer PF, Zhang T. 2013. Widespread production of extracellular superoxide by heterotrophic bacteria. *Science* 340:1223–1226. <https://doi.org/10.1126/science.1237331>.