Draft Genome Sequence from a Putative New Genus and Species in the Family *M1A02* within the Phylum *Planctomycetes*, Isolated from Benthic Pinnacle Mats in Lake Untersee, Antarctica

Nicole Yasmin Wagner,[a] Dale T. Andersen,[b] Aria S. Hahn,[c] Ryan McLaughlin,[c] Sarah Stewart Johnson[a,d]

[a]Department of Biology, Georgetown University, Washington, DC, USA
[b]SETI Institute, Mountain View, California, USA
[c]Koonkie Inc., Menlo Park, California, USA
[d]Science, Technology, and International Affairs Program, Georgetown University, Washington, DC, USA

ABSTRACT Here, we report the draft genome sequence for a new putative genus and species in the family *M1A02* within the order *Phycisphaerales*. Isolated from the metagenome of a benthic pinnacle-shaped mat in the Antarctic Lake Untersee, the members of this family have been found in biofilms and freshwater environments.

Untersee is a perennially ice-covered lake in Antarctica (71°20′31.13″S, 13°27′16.45″E).

Benthic, photosynthetic microbial mats are present to depths exceeding 130 m. Small pinnacle mats up to 15 cm tall form at depths between 10 m and 15 m. Conical structures up to 70 cm tall form at depths between 10 m and 130 m (1, 2). Samples were obtained from pinnacle mats collected at 13 m using 50-mL Falcon tubes in fall 2019. The samples were preserved in 75% ethanol and later stored at −80°C. DNA was extracted using a Qiagen AllPrep kit (Germantown, MD). Library preparation was performed using the Illumina DNA-Prep Workflow (San Diego, CA) and run on the Illumina MiniSeq platform using a MiniSeq reagent midoutput (MO) kit (300 cycles) (Illumina Inc.), generating 151-bp paired-end reads. The reads were trimmed using Trimmomatic v0.39 (3). MetaWRAP v1.3 (4) was used to assemble and bin (using the binning module) the metagenomes into metagenome-assembled genomes (MAGs), with default parameters. In all, 38,626,094 reads were assembled, 3,740,790 of which mapped to the MAG described here.

This MAG was found in a pinnacle mat. It is 94.45% complete and 0.00% contaminated and has a GC content of 67%, as assessed using CheckM v1.1.3 (5) with default parameters. It contains 3,740,790 reads (2.51% of total metagenomic reads) in 59 contigs. An average coverage of 50.8× was calculated using BWA-mem2 v2.2.1 (6) and SAMtools v1.13 (7) with default parameters. The MAG comprises 3,099,537 bp with an N_{50} value of 132,036 bp. There are 2,454 open reading frames (ORFs). The MAG contains 127 tRNA genes and 1 small subunit 16S rRNA gene.

Using GTDB-tk v1.5.0 (8) with default settings, the MAG was classified as belonging to the family *M1A02*, in the order *Phycisphaerales*, phylum *Planctomycetes*. The closest full-length 16S match on NCBI (https://www.ncbi.nlm.nih.gov/), according to a Web-based BLASTN search of the NCBI nonredundant/nucleotide (nr/nt) database (accessed 11 January 2021) (9, 10), was a partial rRNA of an uncultured bacterium discovered in the northern South China Sea, with a 92% identity match (11), below the minimum identity (94.5%) used for 16S genus identification (12), suggesting that this MAG represents a new genus.
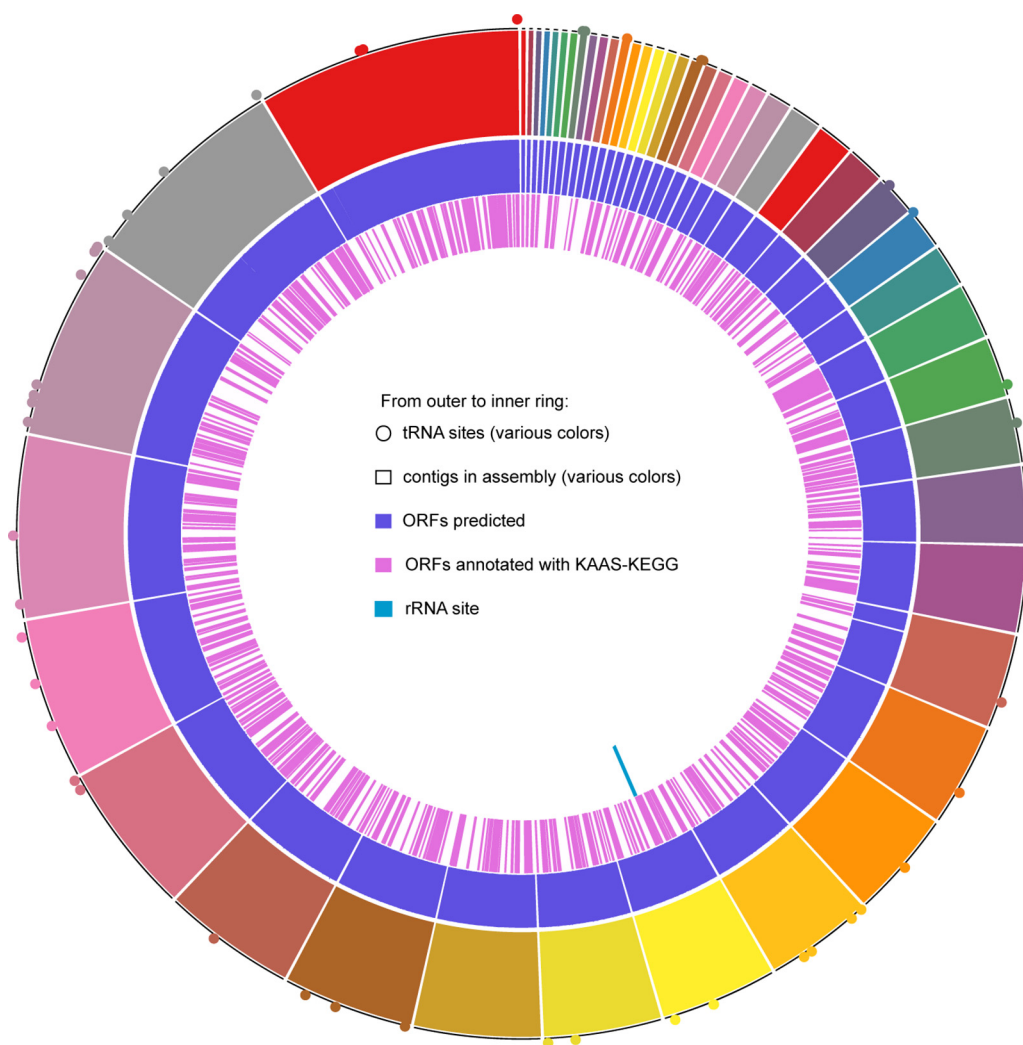
FIG 1 Metagenome-assembled genome (MAG) Circos plot showing the following (from outer to inner rings): first, tRNA locations; contigs based on their order in the MAG; the proportion of total predicted open reading frames (ORFs) (any sequence with a start and stop codon, excluding rRNA genes) with a corresponding annotation in the KEGG database (5); and finally, in the innermost ring, the location of the rRNA.

Functional annotation was performed using KAAS-KEGG v2.1 (13) with default parameters. A total of 18.13% of the ORFs code for nucleotide, amino acid, and lipid metabolism, 23.8% for carbohydrate and energy metabolism, and 32.36% for genetic information processing (Fig. 1). It contains *ppdK* used in the C4 photosynthesis pathway, which is known to increase the efficiency of carbon dioxide fixation in photosynthetic organisms and facilitate ATP production from glycolysis (14, 15). This MAG contains MDH1 and MDH2, present in the C4 photosynthetic pathway (16). Together, this evidence suggests that this new genus is capable of photosynthetic carbon cycling.

We identified ORFs (e.g., *cheR*, *cheB*, *cheA*, *cheW*, *oppA*, *oppB*, and *oppC*) responsible for chemotaxis and quorum-sensing (17), as well as biofilm formation, and a chemosensory system that regulates biofilm formation (*gacA*/*gacS* system and *wspR* [18, 19]). Additionally, we found ORFs related to the formation of flagellum (*fliF*, *fliG*, *fliM*, *fliN*, *flhA*, *flhB*, *motA*, *motB*, and *glfK*), suggesting motility (20, 21), and the presence of quorum-sensing and chemotaxis genes, suggesting that the MAG aids in pinnacle mat formation.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under accession number JAJOLK000000000. This study represents version JAJOLK010000000. The reads were deposited under BioProject accession number

PRJNA784630. The GFF and gene and protein fasta files are accessible at https://doi.org/10.6084/m9.figshare.19376819.v1, and the functional annotation can be found at https://doi.org/10.6084/m9.figshare.19287740.v1.

REFERENCES

1. Andersen DT, Sumner DY, Hawes I, Webster-Brown J, McKay CP. 2011. Discovery of large conical stromatolites in Lake Untersee, Antarctica. Geobiology 9:280–293. https://doi.org/10.1111/j.1472-4669.2011.00279.x.
2. Faucher B, Lacelle D, Marsh NB, Jasperse L, Clark ID, Andersen DT. 2021. Glacial lake outburst floods enhance benthic microbial productivity in perennially ice-covered Lake Untersee (East Antarctica). Commun Earth Environ 2:211. https://doi.org/10.1038/s43247-021-00280-x.
3. 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.
4. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:1–13. https://doi.org/10.1186/s40168-018-0541-1.
5. 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.
6. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997 [q-bio.GN]. https://doi.org/10.48550/arXiv.1303.3997.
7. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
8. 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.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
10. Hu YO, Karlson B, Charvet S, Andersson AF. 2016. Diversity of pico- to mesoplankton along the 2000 km salinity gradient of the Baltic Sea. Front Microbiol 7:679. https://doi.org/10.3389/fmicb.2016.00679.
11. Li J, Li N, Li F, Zou T, Yu S, Wang Y, Qin S, Wang G. 2014. Spatial diversity of bacterioplankton communities in surface water of northern South China Sea. PLoS One 9:e113014. https://doi.org/10.1371/journal.pone.0113014.
12. Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, Emerson DA. 2020. Genus definition for bacteria and archaea based on a standard genome relatedness index. mBio 11:e02475-19. https://doi.org/10.1128/mBio.02475-19.
13. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185. https://doi.org/10.1093/nar/gkm321.
14. Yadav S, Mishra A, Jha B. 2018. Elevated CO_2 leads to carbon sequestration by modulating C_4 photosynthesis pathway enzyme (PPDK) in *Suaeda monoica* and *S. fruticosa*. J Photochem Photobiol B 178:310–315. https://doi.org/10.1016/j.jphotobiol.2017.11.022.
15. Ermakova M, Arrivault S, Giuliani R, Danila F, Alonso-Cantabrana H, Vlad D, Ishihara H, Feil R, Guenther M, Borghi GL, Covshoff S, Ludwig M, Cousins AB, Langdale JA, Kelly S, Lunn JE, Stitt M, Caemmerer S, Furbank RT. 2021. Installation of C_4 photosynthetic pathway enzymes in rice using a single construct. Plant Biotechnol J 19:575–588. https://doi.org/10.1111/pbi.13487.
16. Minges A, Ciupka D, Winkler C, Höppner A, Gohlke H, Groth G. 2017. Structural intermediates and directionality of the swiveling motion of pyruvate phosphate dikinase. Sci Rep 7:1–14. https://doi.org/10.1038/srep45389.
17. Wagner NY, Hahn AS, Andersen D, Wilhelm MB, Morgan-Lang C, Vanderwilt M, Johnson SS. 2019. Draft genome sequence from a putative new genus and species in the family *Methanoregulaceae* isolated from the anoxic basin of Lake Untersee in East Antarctica. Microbiol Resour Announc 8:e00271-19. https://doi.org/10.1128/MRA.00271-19.
18. Parkins MD, Ceri H, Storey DG. 2001. *Pseudomonas aeruginosa* GacA, a factor in multihost virulence, is also essential for biofilm formation. Mol Microbiol 40:1215–1226. https://doi.org/10.1046/j.1365-2958.2001.02469.x.
19. Hickman JW, Tifrea DF, Harwood CS. 2005. A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. Proc Natl Acad Sci U S A 102:14422–14427. https://doi.org/10.1073/pnas.0507170102.
20. Beutin L, Strauch E, Zimmermann S, Kaulfuss S, Schaudinn C, Männel A, Gelderblom HR. 2005. Genetical and functional investigation of fliC genes encoding flagellar serotype H4 in wildtype strains of Escherichia coli and in a laboratory E. coli K-12 strain expressing flagellar antigen type H48. BMC Microbiol 5:4–11. https://doi.org/10.1186/1471-2180-5-4.
21. Sowa Y, Berry RM. 2008. Bacterial flagellar motor. Q Rev Biophys 41:103–132. https://doi.org/10.1017/S0033583508004691.