





Metagenome-Assembled Genome Sequences from an Anoxygenic Photosynthetic Consortium Involved in Sulfur Cycling

Lucia Winkler,^a Marc F. Munker,^a Susanne Brunotte,^a Lina Rohlmann,^a Alvaro Diez Alfageme,^a  Anja Poehlein,^b Michael Hoppert,^c Joachim Reitner,^d  Heiko Nacke^b

^aFoLL Interdisciplinary Students' Research Group, Faculty for Biology and Psychology and Faculty for Geosciences and Geography, Georg-August-Universität Göttingen, Göttingen, Germany

^bDepartment of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen, Göttingen, Germany

^cDepartment of General Microbiology, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen, Göttingen, Germany

^dGeosciences Centre, Department of Geobiology, Georg-August-Universität Göttingen, Göttingen, Germany

ABSTRACT We sequenced the metagenome of an anoxygenic photosynthetic consortium originating from pond water and reconstructed four metagenome-assembled genomes. These genomes include *Desulfocapsa*, *Paludibacter*, *Lamprocystis*, and *Rhodocyclaceae* representatives and indicate the presence of genes for dissimilatory sulfate reduction and oxidation of reduced sulfur compounds.

In the center of the former Willershausen clay pit in Germany (51°47'0.57"N, 10°6'50.85"E), ascending groundwater forms wetlands and ponds, which are hydrogen carbonate-sulfate/sulfate-calcium dominated because of a remolding in the underlying Zechstein salinar (1). Exclusively in the easternmost pond, globular aggregates of pink bacteria ("pink consortia") were found in the water column.

In April 2019, a water sample was taken from the easternmost pond (oxygen content, 1.44 mg/liter; redox potential, -112.1 mV; a Xylem Analytics [Weilheim, Germany] FDO 925 oxygen sensor and an ORP 900 redox electrode were used for these measurements). To remove larger debris and arthropods, the sample was filtered through a stack of sieve pans with pore sizes of 500, 300, 80, and 41 μm (from coarse to fine). Fractions retained by the 80- and 41- μm meshes contained the pink consortia free from the aforementioned contaminants. DNA was extracted using the DNeasy PowerSoil kit according to the manufacturer's protocol (Qiagen, Hilden, Germany). The purified DNA was used to generate Illumina paired-end sequencing libraries with the Nextera DNA sample preparation kit (Illumina, San Diego, CA, USA), and the libraries were sequenced by employing the MiSeq reagent kit v.3 and a MiSeq instrument, as stated by the manufacturer (Illumina). Default parameters were used for all software unless otherwise specified. Quality trimming of reads was performed by applying the program fastp v.0.19.4 (qualified quality Phred score, 20; minimum read length, 50 bp) (2) and yielded 26,465,778 paired-end reads. An average read length of 245 bp was recorded for the forward reads, and an average read length of 239 bp was recorded for the reverse reads. Base correction in overlapping regions (the correction option was selected with respect to fastp-based quality trimming using default parameters; this option allows identification of overlapping regions of each pair of reads, and mismatched base pairs in these regions can be corrected if one base shows high quality and the other very low quality) and removal of automatically detected adapter sequences were performed. Low-quality bases at the 5' and 3' ends of reads were trimmed once the mean quality within a sliding window of 4 dropped below 20. Sequences were *de novo* assembled into 5,933 contigs of $\geq 1,000$ bp by utilizing

Citation Winkler L, Munker MF, Brunotte S, Rohlmann L, Diez Alfageme A, Poehlein A, Hoppert M, Reitner J, Nacke H. 2020. Metagenome-assembled genome sequences from an anoxygenic photosynthetic consortium involved in sulfur cycling. *Microbiol Resour Announc* 9:e00819-20. <https://doi.org/10.1128/MRA.00819-20>.

Editor David A. Baltrus, University of Arizona

Copyright © 2020 Winkler 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 Heiko Nacke, hacke@gwdg.de.

Received 21 July 2020

Accepted 5 September 2020

Published 1 October 2020

metaSPAdes v.3.13.1 (3). The fastp and metaSPAdes outputs were fed into MaxBin v.2.2.7 for binning (minimum contig length, 1,000 bp; minimum probability for binning, 0.50) (4). Application of CheckM v.1.0.18 (5) revealed four relatively complete metagenome-assembled genomes (MAGs) (completeness, $\geq 94\%$; contamination rate, $\leq 2.1\%$). These four MAGs were annotated by using Prokka v.1.14.5 (6). Subsequently, the Prokka output was analyzed using the PathoLogic (7) component of the Pathway Tools software v.23.5 (8) with the MetaCyc database v.23.5 (9). Taxonomic analysis based on 16S rRNA gene sequences was performed using the Type (Strain) Genome Server (TYGS) v.205 (10). In addition, MAGs were classified taxonomically using GTDB-Tk v.1.0.2 and the Genome Taxonomy Database (GTDB) (release 89) (11, 12).

According to TYGS-based 16S rRNA gene analysis, MAG 01 (length of detected 16S rRNA gene sequence, 1,546 bp) is closely related to *Desulfocapsa sulfexigens* and MAG 03 (length of detected 16S rRNA gene sequence, 1,514 bp) to *Paludibacter propionigenes* (with respect to MAG 02 and MAG 04, no 16S rRNA gene sequences were detected). GTDB-Tk-based classification confirmed that MAG 01 and MAG 03 are affiliated with *Desulfocapsaceae* and *Paludibacter*, respectively, and revealed that MAG 02 belongs to *Lamprocystis* and MAG 04 belongs to *Rhodocyclaceae*. Functional analysis revealed the presence of genes for dissimilatory sulfate reduction, encoding, e.g., dissimilatory adenylyl-sulfate reductases (AprA and AprB) and sulfite reductases (DsvA, DsvB, DsvC, and AsrA), in MAGs 01, 02, and 04. All four MAGs encode the polysulfide reductase NrfD, indicating potential polysulfide formation from sulfur. MAG 02 and MAG 04 harbor genes for oxidation of reduced sulfur compounds, e.g., genes encoding the sulfide dehydrogenase subunits FccA and FccB and the sulfite dehydrogenase SoeB. Furthermore, both MAGs contain genes encoding parts of a Sox system (*soxA* and *soxY*).

Data availability. Raw sequencing data are available at the NCBI Sequence Read Archive (SRA) under accession number [SRR11932628](https://www.ncbi.nlm.nih.gov/sra/SRR11932628). The metagenome assembly is available at GenBank under accession number [JACDZJ000000000](https://www.ncbi.nlm.nih.gov/genbank/JACDZJ000000000). The MAGs are available at GenBank under accession numbers [JACDZF000000000](https://www.ncbi.nlm.nih.gov/genbank/JACDZF000000000), [JACDZG000000000](https://www.ncbi.nlm.nih.gov/genbank/JACDZG000000000), [JACDZH000000000](https://www.ncbi.nlm.nih.gov/genbank/JACDZH000000000), and [JACDZI000000000](https://www.ncbi.nlm.nih.gov/genbank/JACDZI000000000). Prokka-based annotations of contigs with respect to the four MAGs are publicly available at the Göttingen Research Online Database (<https://doi.org/10.25625/EZNVW8>).

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research within the context of Göttingen Campus QPLUS. We gratefully acknowledge the generous support of the Heimatverein Willershausen e.V. and the FoLL project of the section for Teaching and Learning in Higher Education at Georg-August-Universität Göttingen. We also acknowledge support by the Open Access Publication Funds of the University of Göttingen. The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

We thank Melanie Heinemann and Sarah Teresa Schüßler for technical support.

REFERENCES

1. Meischner D. 2000. Der pliozäne Teich von Willershausen am Harz, p 223–228. In Meischner D (ed), *Europäische fossillagerstätten*. Springer, Berlin, Germany.
2. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
3. 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>.
4. Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2:26. <https://doi.org/10.1186/2049-2618-2-26>.
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. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
7. Karp PD, Latendresse M, Caspi R. 2011. The Pathway Tools pathway prediction algorithm. *Stand Genomic Sci* 5:424–429. <https://doi.org/10.4056/signs.1794338>.
8. Karp PD, Latendresse M, Paley SM, Krummenacker M, Ong QD, Billington R, Kothari A, Weaver D, Lee T, Subhraveti P, Spaulding A, Fulcher R, Keseler IM, Caspi R. 2016. Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform* 17:877–890. <https://doi.org/10.1093/bib/bbv079>.
9. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. 2020. The MetaCyc database of metabolic pathways and enzymes: a 2019 update. *Nucleic Acids Res* 48:D445–D453. <https://doi.org/10.1093/nar/gkz862>.

10. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>.
11. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2019. Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. *bioRxiv* <https://doi.org/10.1101/771964>.
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>.