# Complete Genome Sequence of an Uncultivated Freshwater Bacteroidota Lineage

Emanuel Mauch,[a] Lucas Serra Moncadas,[b] Adrian-Stefan Andrei[b]

aFaculty of Science, University of Zurich, Zurich, Switzerland

bMicrobial Evogenomics Lab, Limnological Station, Department of Plant and Microbial Biology, University of Zurich, Kilchberg, Switzerland

**ABSTRACT** We report here a complete metagenome-assembled genome belonging to the AKYH767 order within the Bacteroidota phylum. The recovered genome stems from a nonaxenic Amoebozoa culture from Lake Zürich and was assembled as a circular element at a length of 4.1 Mbp and a coverage of 115$\times$.

Bacteroidota phylum members have colonized a large spectrum of ecological niches, including soil, freshwater, marine and animal-associated, displaying rich metabolic potential. Bacteroidota emerged as a major freshwater bacterial lineage in 16S rRNA sequences analyses recovered from diverse aquatic habitats, playing a crucial role in degrading complex organic matter (1).

Unidentified Amoebozoa was nonaxenically cultured from a Lake Zürich (47° 13′ 20.9994″N, 8° 45′ 9.72″E, Swiss Confederation) sample. The cultures were maintained in tissue culture flasks filled with 50 mL of autoclaved Swiss Alpina water (Pearlwater, Termen, Switzerland). *Planktothrix rubescens* strain A7 served as food source. The DNA of this culture was extracted using the MagAttract HMW DNA kit (Qiagen, Hilden, Germany). It was further purified with Beckman Coulter AMPure XP magnetic beads and subsequently used for metagenomic sequencing on a Nanopore PromethION platform using a FLO-PRO002 (R9.4.1) flow cell. The sequencing 1D library was constructed with the SQK-LSK110 Ligation Sequencing kit (ONT, Oxford, UK) in conformity with manufacturer's instructions without any prior DNA fragmentation or size selection.

Obtained reads (approx. 2 million reads; mean read length 9 kbp, median read quality 11.7, $N_{50}$ 14.4 kbp) were basecalled and quality trimmed (Q score 7) with Guppy 5.1.15 prior to assembly with Flye 2.9-b1768 (2). Potential contamination of the recovered circular chromosome (4 155 722 bp, coverage 115$\times$) was assessed by CheckM v1.1.3 (94.89% complete, 0.63% contaminated) (3). Taxonomical classification was performed with GTDB-Tk v1.4.0 (4) and by comparing its 16S rRNA, predicted by barrnap 0.9, against SILVA database v138 (5). Prokka 1.12 (6) and NCBI's PGAP pipeline were used for gene prediction and functional annotation. The assignment of KO identifiers to orthologous genes was performed by BlastKOALA (7). Pathway reconstruction was accomplished with the online KEGG Mapper-Reconstruction tool using the obtained KO numbers. PFAM domains were predicted using the pfam_scan.pl script with the PFAM database release 32 (8). gRodon R package (9) was applied to predict growth rates, while defense mechanisms diversity was explored with DefenseFinder (10). Default parameters were used except where otherwise specified.

GTDB-tk classified the obtained metagenome-assembled genome as belonging to an uncultivated order (AKYH767) within the Bacteroidota phylum. However, SILVA classification pointed toward Sphingobacteriales (88.45% identity) order within the same phylum. With a GC of 40.07%, the recovered metagenome-assembled genome exhibited four 869 CDS, 70 encoded transporters, and one rRNA operon. Genome-inferred metabolic reconstruction suggested a Gram-negative bacterium with fast duplication time (approx. 2.6 h) and an aerobic heterotrophic lifestyle. The circular chromosome (strain JAD_PAG50586_3) had the metabolic

potential of synthesizing 50% of all proteinogenic amino acid as well as to degrade 60% of them. Anabolic (i.e., Gluconeogenesis, fatty acid synthesis) and catabolic (i.e., TCA-cycle, pyruvate oxidation, fatty acids degradation) metabolisms were found complete. Interestingly, the recovered genome displayed seven different defense mechanisms. Gliding motility was inferred based on the presence of 11 key genes.

**Data availability.** All sequence data are available through the National Center for Biotechnology Information (NCBI) via the BioProject PRJNA824509 (CP096135.1; GCA_023213195.1; SRX14779367). Additional proteome annotations (KEGG, Prokka and Pfam) are available in figshare repository: 10.6084/m9.figshare.20318880.

## REFERENCES

1. Thomas F, Hehemann JH, Rebuffet E, Czjzek M, Michel G. 2011. Environmental and gut Bacteroidetes: the food connection. Front Microbiol 2:1–16. https://doi.org/10.3389/fmicb.2011.00093.

2. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17:1103–1110. https://doi.org/10.1038/s41592-020-00971-x.

3. 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.

4. Chaumeil PA, 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.

5. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.

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

7. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.

8. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. 2014. Pfam: the protein families database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223.

9. Weissman JL, Hou S, Fuhrman JA. 2021. Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. Proc Natl Acad Sci USA 118:1–10. https://doi.org/10.1073/pnas.2016810118.

10. Tesson F, Hervé A, Touchon M, D'humières C, Cury J, Bernheim A. 2022. Systematic and quantitative view of the antiviral arsenal of prokaryotes. Nat Commun 13:2561. https://doi.org/10.1038/s41467-022-30269-9.