A high-quality genome of an undescribed Flavobacterium species uncovered using Q20+ Nanopore chemistry

Hofer, Cyrill ; Wälchli, Denise ; Moncadas, Lucas Serra ; Rain-Franco, Angel ; Andrei, Adrian-Stefan

DOI: https://doi.org/10.1128/mra.00716-23

Microbiology Resource Announcements

| Computational Biology | Announcement

A high-quality genome of an undescribed *Flavobacterium* species uncovered using Q20+ Nanopore chemistry

Cyrill Hofer,[1] Denise Wälchli,[2] Lucas Serra Moncadas,[1] Angel Rain-Franco,[1] Adrian-Stefan Andrei[1]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Herein, we document the complete genome of the *Flavobacterium* strain ZE23DGlu08, isolated from Lake Zurich, Switzerland. The circular genome was assembled using long-read Nanopore data (coverage: 226×) with the Q20+ chemistry. The described strain displays a genome size of ~3.9 Mbp with a GC content of 34%.

KEYWORDS genomics, DNA sequencing, Nanopore, *Flavobacterium*

While some representatives of the genus *Flavobacterium* are known as severe fish pathogens (1), the vast majority exhibit a free-living lifestyle in wet soils, rivers, lakes, and oceans (2). Primarily, Flavobacteria are aerobic and chemo-organotrophic organisms specialized to degrade high-molecular-weight-dissolved organic compounds, thereby contributing to the carbon cycle of aquatic ecosystems (3).

The strain was isolated from the surface water of Lake Zurich (47°18′N, 8°34′E, Switzerland) and cultivated in agar plates (15 g/L) amended with artificial lake water with cellobiose as carbon substrate (1 mM final concentration), following a published protocol (4). After 21 days of cultivation, colonies were purified through three sequential streaking cycles and then cryopreserved (5). The isolated strain was reactivated in artificial lake water (4), supplemented with glucose (100 µM) for 1 week, and grown in commercial LB medium (Difco, 240210) for 2 days. DNA extraction was performed using the Quick-DNA HMW MagBead Kit (Zymo Research) and cleaned with Agencourt AMPure XP beads (Beckman Coulter) according to the manufacturer's guidelines. Library preparation was done using the Native Barcoding Kit 24 V14 (SQK-NBD114-24, Oxford Nanopore) by processing 1 µg of pure (A260/280 = 1.824, A260/230 = 2.252) high molecular weight DNA (~50 kb, GeneRuler High Range DNA Ladder—Thermo Scientific). Without additional shearing, 154.5 ng was loaded on a FLO-MIN114 R10.4.1 flow cell and sequenced on the minION Mk1B (Oxford Nanopore). After 72 h, the sequencing yielded 1,272,573 reads passing the quality score of Q10 (mean quality: Q14.2; N50: 6,610 bp) before further filtration using chopper [v.0.2.0 (6)] to retain high-quality reads only (≥Q17, ≥1,000 bp). The raw data were basecalled using guppy [v.6.4.6 (7); super accurate basecalling, 400 bps]. The obtained file (177,273 reads; mean quality: Q18.1; N50: 6,819 bp) was *de novo* assembled using the Flye assembler [v.2.9.2-b1786 (8)] by specifying the corrected Nanopore data option, supported by the high read quality with less than 3% average error (flye --nano-hq –meta). One circular contig (identified and trimmed by Flye) was recovered with a length of 3,893,827 bp and a GC content of 34% (99.29% completeness, 0.24% contamination). The quality of the assembled sequence was assessed by CheckM [v.1.2.2 (9)], before whole-genome taxonomic classification with GTDB-Tk [v.2.2.6 (10)] against the Genome Taxonomy Database 214.1 (11). The 16S rRNA genes were extracted using barrnap [v.0.9 (12)] and classified with SILVA 138.1 SSU (13) and NCBI (14). Gene prediction was performed using Prokka [v.1.12 (15)] as well as NCBI's PGAP [v.6.5 (16)] pipeline, followed by annotation, harnessing the PFAM (17), KEGG (18),

Editor Irene L. G. Newton, Indiana University, Bloomington, Bloomington, Indiana, USA

Address correspondence to Cyrill Hofer, cyrill.hofer@uzh.ch, or Adrian-Stefan Andrei, stefan.andrei@limnol.uzh.ch.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 4 August 2023
Accepted 3 November 2023
Published 1 December 2023

MetaCyc (19), and NCBI (14) databases. Default parameters were used except where otherwise noted.

The isolated strain was classified as an uncultured *Flavobacterium* species displaying 98.15% (coverage: 90%) whole-genome similarity with a genome recovered from freshwater sediments (i.e., GCF_002754315.1). The closest described species was found to be *Flavobacterium franklandianum* [NCBI (14): MK346164.1] sharing a 98.55% 16S rRNA identity (coverage: 100%), despite the low-genome similarity of 77.84% (coverage: 42.61%). The genome-resolved functional annotation of the 3,355 genes revealed four intact rRNA operons, such as complete amino acid biosynthesis, glycolysis and glycogenesis, the presence of the TCA cycle, and oxidative phosphorylation. In conclusion, this genome belongs to a copiotrophic *Flavobacterium* species that exhibits a non-motile, free-living, aerobic, and heterotrophic lifestyle.

ACKNOWLEDGMENTS

This study has been completed throughout the course BIO260: Aquatic Microbial Ecology, hosted by members of the Limnological Station, University of Zurich.

This work was supported by the Ambizione grant PZ00P3193240 (Swiss National Science Foundation).

Special thanks goes to Bettina Sieber for isolating this strain.

AUTHOR AFFILIATIONS

[1]Limnological Station, Department of Plant and Microbial Biology, University of Zurich, Kilchberg, Switzerland
[2]Faculty of Science, University of Zurich, Zurich, Switzerland

AUTHOR ORCIDs

Cyrill Hofer http://orcid.org/0000-0002-1516-7626
Angel Rain-Franco http://orcid.org/0000-0002-0389-7154
Adrian-Stefan Andrei http://orcid.org/0000-0003-1425-7168

FUNDING

Funder	Grant(s)	Author(s)
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	193240	Cyrill Hofer
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	193240	Adrian-Stefan Andrei
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNF)	193240	Lucas Serra Moncadas

AUTHOR CONTRIBUTIONS

Cyrill Hofer, Software, Supervision, Writing – original draft, Writing – review and editing | Denise Wälchli, Formal analysis, Writing – original draft | Lucas Serra Moncadas, Supervision, Writing – original draft, Writing – review and editing | Angel Rain-Franco, Methodology, Writing – review and editing | Adrian-Stefan Andrei, Conceptualization, Project administration, Supervision, Validation, Writing – review and editing

DATA AVAILABILITY

The sequence data and additional information are available at NCBI, BioProject PRJNA991734 (Accession: CP130045, BioSample: SAMN36315993, SRA: SRR25177534).

REFERENCES

1. Bernardet JF, Segers P, Vancanneyt M, Berthe F, Kersters K, Vandamme P. 1996. Cutting a gordian knot: emended classification and description of the genus *Flavobacterium*, emended description of the family flavobacteriaceae, and proposal of *Flavobacterium* hydatis nom. Int j syst bacteriol 46:128–148. https://doi.org/10.1099/00207713-46-1-128
2. Bernardet JF, Nakagawa Y. 2006. An introduction to the family *Flavobacteriaceae*, p 455–480. Springer New York, New York, NY.
3. Kirchman DL. 2002. The ecology of cytophaga-flavobacteria in aquatic environments. FEMS Microbiol Ecol 39:91–100. https://doi.org/10.1111/j.1574-6941.2002.tb00910.x
4. Le Moigne A, Randegger F, Gupta A, Petchey OL, Pernthaler J. 2023. Stochasticity causes high β-diversity and functional divergence of bacterial assemblages in closed systems. Ecology 104:e4005. https://doi.org/10.1002/ecy.4005
5. Rain-Franco A, de Moraes GP, Beier S. 2020. Cryopreservation and resuscitation of natural aquatic prokaryotic communities. Front Microbiol 11:597653. https://doi.org/10.3389/fmicb.2020.597653
6. De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics 39:btad311. https://doi.org/10.1093/bioinformatics/btad311
7. Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford nanopore sequencing. Genome Biol 20:129. https://doi.org/10.1186/s13059-019-1727-y
8. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8
9. 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
10. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics 38:5315–5316. https://doi.org/10.1093/bioinformatics/btac672
11. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for bacteria and Archaea. Nat Biotechnol 38:1098. https://doi.org/10.1038/s41587-020-0539-7
12. Torsten S. 2023. Barrnap 0.9: rapid ribosomal RNA prediction. Available from: https://github.com/tseemann/barrnap
13. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and "all-species living tree project (LTP)" Taxonomic frameworks. Nucleic Acids Res 42:D643–D648. https://doi.org/10.1093/nar/gkt1209
14. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST. 2022. Database resources of the national center for biotechnology information. Nucleic Acids Res 50:D20–D26. https://doi.org/10.1093/nar/gkab1112
15. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
16. 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
17. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. Nucleic Acids Res 49:D412–D419. https://doi.org/10.1093/nar/gkaa913
18. Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27
19. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, Ong WK, Paley SM, Subhraveti P. 2019. The BioCyc collection of microbial genomes and metabolic pathways. Brief Bioinform 20:1085–1093. https://doi.org/10.1093/bib/bbx085