AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Metagenome-Assembled Genomes for "*Candidatus* Phormidium sp. Strain AB48" and Co-occurring Microorganisms from an Industrial Photobioreactor Environment

Avery J. C. Noonan,[a,b] Yilin Qiu,[a] Brandon Kieft,[c] Sean Formby,[d] Tony Liu,[d] Kalen Dofher,[a] Moritz Koch,[c] Steven J. Hallam[a,b,c,d,e]

[a]Genome Science and Technology Program, University of British Columbia, Vancouver, British Columbia, Canada
[b]ECOSCOPE Training Program, University of British Columbia, Vancouver, British Columbia, Canada
[c]Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada
[d]Graduate Program in Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada
[e]Life Sciences Institute, University of British Columbia, Vancouver, British Columbia, Canada

Avery J. C. Noonan and Yilin Qiu contributed equally to this work. Author order was determined alphabetically.

ABSTRACT Here, we report metagenome-assembled genomes for "*Candidatus* Phormidium sp. strain AB48" and three cooccurring microorganisms from a biofilm-forming industrial photobioreactor environment, using the PacBio sequencing platform. Several mobile genetic elements, including a double-stranded DNA phage and plasmids, were also recovered, with the potential to mediate gene transfer within the biofilm community.

Cyanobacteria offer the possibility of producing energy and materials from sunlight and carbon dioxide (1, 2). However, these applications of cyanobacteria remain limited, and more industrial strains with desirable growth properties are needed (1, 3, 4). Recent studies indicate that members of the *Phormidium* genus have potential to fill this gap (5–7). The genus represents a polyphyletic distribution of filamentous cyanobacteria containing over 200 species, many of which form dense biofilms with co-occurring microorganisms (8–10).

Here, we describe metagenome-assembled genome (MAG) sequences, including mobile genetic elements (MGEs), from an industrial photobioreactor environment. The photobioreactor uses high temperature (35 to 45°C), salinity (10 g/L), and alkaline conditions (pH 9 to 11) to support biofilm-based growth. After several months of continuous cultivation, with harvesting every 48 to 96 h, a biofilm sample was collected. Genomic DNA was extracted from biofilm biomass using a cetyltrimethylammonium bromide (CTAB)-chloroform extraction protocol (11). DNA was sheared to 10 kb using Covaris g-TUBES and size selected (>10 kb) with AMPure beads (Beckman Coulter). This high-molecular-weight fraction was used to prepare a Pacific Biosciences (PacBio) SMRTbell library for sequencing on the PacBio Sequel platform. Default parameters were used for sequencing data processing and analysis software, unless otherwise noted. A total of 1,551,061 reads (5,417 Mbp) were filtered and trimmed using BBTools (v.38.79) (parameters: jni=t json=t ow=t cq=f keepshortreads=f trim=f), resulting in 1,417,931 reads (4,615 Mbp) with an N_{50} value of 5,590 bp. The metagenome was assembled with Flye (v.2.7) using meta parameters, and the Flye assembly graph was converted into FASTA format for downstream processing (12). Contigs of >1,000 bp were assigned to population genome bins using MaxBin2 (v.2.2.6) (13), and the completeness and contamination of resulting MAGs were assessed using CheckM (v.1.1.3) (14).

A total of 281 contigs, with an N_{50} value of 56,161 bp and a total length of 13,019,354 bp as evaluated by QUAST (v.5.0.2) (15), were assembled. Four MAGs were resolved, including a complete genome and three additional population genome bins of varying quality. The complete genome was 4,818,683 bp, with a GC content of 51.64% and an N_{50} value of 4,751,363 bp. A high-quality genome of 3,207,041 bp, with CheckM completeness of 89.03%,

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

Address correspondence to Steven J. Hallam, shallam@mail.ubc.ca.

The authors declare a conflict of interest.

Received 3 August 2022
Accepted 27 October 2022
Published 21 November 2022

a GC content of 49.23%, and an N_{50} value of 51,694 bp, and two draft-quality genomes of 3,511,511 bp and 1,359,261 bp, with completeness values of 37.80% and 23.31%, GC contents of 59.57% and 63.91%, and N_{50} values of 30,355 bp and 25,610 bp, respectively, were identified. MGEs, including putative phages and plasmids, were identified using PlasFlow (v.1.1), VirSorter2 (v.2.2.2), and CheckV (v.0.8.1) (16–18). A total of three circular plasmids, one high-quality double-stranded DNA (dsDNA) phage (100% completeness), and one medium-quality dsDNA phage (50.59% completeness) were predicted. One of the plasmids, 51,598 bp in length, clustered with the complete cyanobacterial MAG. The remaining contigs passing the quality control (QC) threshold (0.94% of total bases) could not be assigned.

Analysis of single-copy marker genes in the four MAGs using GTDB-Tk (v.0.3.2) (19) and an average nucleotide identity (ANI) of 87.75% with its closest relative, *Phormidium_A* sp007126595 (GCA_007131565.1), as calculated using fastANI (v.1.1) (20), indicated that the complete genome represents a novel member of the cyanobacterial *Phormidium_A* genus, designated "*Candidatus* Phormidium sp. strain AB48" (21). The high-quality genome was assigned to the *Alteromonadaceae* family, and the two draft-quality genomes were assigned to the *Verruco 01* and *Maricaulaceae* families. Open reading frame (ORF) prediction and genome annotations were performed using Prokka (v.1.14.5) (22). The complete "*Candidatus* Phormidium sp. strain AB48" genome contains 4,372 genes, 4,318 coding sequences (CDSs), 6 rRNA genes (two 5S rRNAs, two 16S rRNAs, and two 23S rRNAs), and 47 tRNAs.

Data availability. Data from this project are publicly accessible through the NCBI under BioProject accession no. PRJNA834472 and BioSample accession no. SAMN28044958. Raw sequencing data are available from the Sequence Read Archive (SRA) under the SRA accession no. SRR13132258. This BioProject also includes BioSample SAMN28044976; this represents a distinct sample and contains Oxford Nanopore Technologies and Illumina data sets for *Phormidium yuhuli* AB48, which was isolated from the photobioreactor enrichment described here (23).

ACKNOWLEDGMENTS

We thank Soheyl Mottahedeh, Manisha Shastri, Craig Fourie, and Kevin Wilson at AlgaBloom for providing photobioreactor biomass, as well as Tanja Woyke at the Joint Genome Institute (JGI) and Tom Pfeifer in the Biofactorial Automation Core Facility at the University of British Columbia for technical advice and support.

This work was performed under the auspices of the Natural Sciences and Engineering Research Council (NSERC) of Canada, Genome British Columbia, the Canada Foundation for Innovation (CFI), the G. Unger Vetlesen and Ambrose Monell Foundations, the U.S. Department of Energy (DOE) JGI, and the Facilities Integrating Collaborations for User Science (FICUS) JGI Environmental Molecular Science Laboratory (EMSL) (project 50967) supported by the Office of Science of the U.S. DOE (contract DE-AC02-05CH11231), with essential automation support through the Biofactorial Automation Core Facility in the Life Sciences Institute at the University of British Columbia. A.J.C.N. was supported by the NSERC CREATE Ecosystem Services, Commercialization Platforms, and Entrepreneurship (ECOSCOPE) training program at the University of British Columbia and the Mitacs Global Link program. Y.Q., A.J.C.N., and K.D. were also supported by the NSERC CREATE Genome Science and Technology (GSAT) training program at the University of British Columbia.

S.J.H. is a cofounder of Koonkie Inc., a bioinformatics consulting company that designs and provides scalable algorithmic and data analytics solutions in the cloud.

REFERENCES

1. Berla BM, Saha R, Immethun CM, Maranas CD, Moon TS, Pakrasi HB. 2013. Synthetic biology of cyanobacteria: unique challenges and opportunities. Front Microbiol 4:246. https://doi.org/10.3389/fmicb.2013.00246.
2. Nielsen J, Lee S, Stephanopoulos G, Hudson P (ed). 2021. Cyanobacteria biotechnology. Wiley VCH, Weinheim, Germany. https://doi.org/10.1002/9783527824908.
3. Price S, Kuzhiumparambil U, Pernice M, Ralph P. 2022. Techno-economic analysis of cyanobacterial PHB bioplastic production. J Environ Chem Eng 10:107502. https://doi.org/10.1016/j.jece.2022.107502.
4. Taton A, Ecker A, Diaz B, Moss NA, Anderson B, Reher R, Leão TF, Simkovsky R, Dorrestein PC, Gerwick L, Gerwick WH, Golden JW. 2020. Heterologous expression of cryptomaldamide in a cyanobacterial host. ACS Synth Biol 9:3364–3376. https://doi.org/10.1021/acssynbio.0c00431.
5. Ataeian M, Vadlamani A, Haines M, Mosier D, Dong X, Kleiner M, Strous M, Hawley AK. 2021. Proteome and strain analysis of cyanobacterium *Candidatus* "Phormidium alkaliphilum" reveals traits for success in biotechnology. iScience 24:103405. https://doi.org/10.1016/j.isci.2021.103405.
6. Nies F, Mielke M, Pochert J, Lamparter T. 2020. Natural transformation of the filamentous cyanobacterium *Phormidium lacuna*. PLoS One 15:e0234440. https://doi.org/10.1371/journal.pone.0234440.
7. Nies F, Wörner S, Wunsch N, Armant O, Sharma V, Hesselschwerdt A, Falk

F, Weber N, Weiß J, Trautmann A, Posten C, Prakash T, Lamparter T. 2017. Characterization of *Phormidium lacuna* strains from the North Sea and the Mediterranean Sea for biotechnological applications. Process Biochem 59:194–206. https://doi.org/10.1016/j.procbio.2017.05.015.
8. Palinska KA, Deventer B, Hariri K, Łotocka M. 2011. A taxonomic study on *Phormidium*-group (cyanobacteria) based on morphology, pigments, RAPD molecular markers and RFLP analysis of the 16S rRNA gene fragment. Fottea 11:41–55. https://doi.org/10.5507/fot.2011.006.
9. Palinska KA, Marquardt J. 2008. Genotypic and phenotypic analysis of strains assigned to the widespread cyanobacterial morphospecies *Phormidium autumnale* (Oscillatoriales). Arch Microbiol 189:325–335. https://doi.org/10.1007/s00203-007-0323-9.
10. Sciuto K, Andreoli C, Rascio N, La Rocca N, Moro I. 2012. Polyphasic approach and typification of selected *Phormidium* strains (Cyanobacteria). Cladistics 28:357–374. https://doi.org/10.1111/j.1096-0031.2011.00386.x.
11. Morin N, Vallaeys T, Hendrickx L, Natalie L, Wilmotte A. 2010. An efficient DNA isolation protocol for filamentous cyanobacteria of the genus *Arthrospira*. J Microbiol Methods 80:148–154. https://doi.org/10.1016/j.mimet.2009.11.012.
12. 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.
13. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.
14. 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.
15. 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.
16. Krawczyk PS, Lipinski L, Dziembowski A. 2018. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Res 46:e35. https://doi.org/10.1093/nar/gkx1321.
17. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. Microbiome 9:37. https://doi.org/10.1186/s40168-020-00990-y.
18. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol 39:578–585. https://doi.org/10.1038/s41587-020-00774-7.
19. 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.
20. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.
21. Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A, DiCuccio M. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. Int J Syst Evol Microbiol 68:2386–2392. https://doi.org/10.1099/ijsem.0.002809.
22. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
23. Koch M, Noonan A, Qiu Y, Dofher K, Kieft B, Mottahedeh S, Shastri M, Hallam SJ. 2022. The survivor strain: isolation and characterization of Phormidium yuhuli AB48, a filamentous phototactic cyanobacterium with biotechnological potential. Frontiers in bioengineering and biotechnology, 10:932695. https://doi.org/10.3389/fbioe.2022.932695.