Metagenome-assembled genome sequences of two cyanobacterial cultures from Homa Bay County, Kenya

Katelyn M. Brown,[1,2] Christopher S. Ward,[1,2] George S. Bullerjahn[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Metagenome-assembled genomes were generated for two xenic cyanobacterial strains collected from aquatic sources in Kenya and sequenced by NovaSeq S4. Here, we report the classification and genome statistics of *Microcystis panniformis* WG22 and *Limnospira fusiformis* LS22.

Microcystis spp. are cosmopolitan cyanobacteria that tolerate a wide range of temperature conditions, blooming at a minimum of 15°C (1). *Arthrospira* (*Limnospira*) spp. are primarily found in Africa, Asia, South America, and Central America, and occur in soda lakes in the African Rift Valley (2). In June 2022, a *Microcystis* strain was collected from the Winam Gulf offshore of Homa Bay (−0.494583, 34.444167) using a plankton net. Also in June 2022, a *Limnospira* strain was collected in Homa Bay County from Lake Simbi (−0.367750, 34.629833) from the shoreline. Xenic cyanobacterial cultures were separated through a dilution series by selectively pipetting colonies/filaments from 10 µL of original sample on a microscope slide into 10 µL of BG-11 media, repeating until single colonies/filaments were achieved. The resulting *Microcystis* colony was transferred into fresh liquid BG-11 media (https://utex.org/products/bg-11-medium?variant=30991786868826#recipe) in a 24-well plate and incubated at 21°C and 5 µmol/m^2/s until biomass accumulated. *Limnospira* was grown in liquid Spirulina media (pH 10.4; https://utex.org/products/spirulina-medium?variant=30991737454682#recipe). Unialgal growth was confirmed by microscopy, and biomass was transferred into 25 mL of fresh media in a culturing flask.

The xenic cyanobacterial cultures were cultured at the conditions stated above and monitored for growth. After several months of acclimation, approximately 20 mL of dense culture was filtered through a Sterivex filtration unit (0.22 µm pore size, Sigma Aldrich, St. Louis, MO). Filters were frozen until extraction, where the membranes were removed from the plastic casing, and DNA was extracted using a DNeasy PowerWater Kit (Qiagen, Germantown, MD) according to manufacturer's instructions. Eluted DNA was sequenced at the University of Minnesota Genomics Center. Unique Dual Indexed Illumina DNA libraries were prepared using Nextera DNA Flex and sequenced on a NovaSeq S4 to generate 150-bp paired-end metagenomic reads. Paired-end reads were input into the Department of Energy Systems Biology Knowledgebase for *de novo* assembly of each metagenome-assembled genome (MAG) in separate workflows [KBase; (3)]. Default parameters were used for all applications listed unless otherwise specified. Reads were imported as a paired-end library and trimmed to eliminate low-quality base calls and Nextera-PE sequencing adapters, also setting the head crop length to 15 [Trimmomatic v0.36; (4)]. After trimming, read quality was assessed using FastQC (v0.11.9; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Metagenomic reads were assembled using metaSPAdes [v3.15.3; (5)] and binned with MaxBin2 [v2.2.4; (6)], MetaBAT2 [v1.7; (7)], and CONCOCT [v1.1; (8)] with minimum

TABLE 1 Summary of metagenome-assembled genomes WG22 and LS22

Characteristic	*Microcystis panniformis* WG22	*Limnospira fusiformis* LS22
Assembly GenBank accession no.	SAMN37196505	SAMN37196506
Raw reads GenBank accession no.	SAMN36615257	SAMN36615258
No. of reads	408,836,714	506,895,224
MAG length (bp)	4,264,909	5,209,155
Bin completeness (%)	92.12	98.18
Bin contamination (%)	3.51	0.44
No. of reference genomes for marker sets	79	79
No. of marker genes	582	584
No. of marker sets	456	458
Marker genes identified zero times	46	10
Marker genes identified one time	517	571
Marker genes identified two times	19	3
Average GC content (%)	42.71	44.5
No. of contigs	532	407
N_{50} (bp)	13,379	20,557
No. of predicted genes	4,170	4,801
FastANI reference	GCF_010196425.1	GCA_012516315.1
FastANI reference identity	*Microcystis panniformis*	*Limnospira fusiformis*
FastANI ANI (%)	95.6	99.38

contig length of 2,000 bp prior to bin optimization using DAS-Tool [v1.1.2; (9)]. Bin quality was assessed with CheckM [v1.0.18; Table 1; (10)], and the bins classified to cyanobacteria were extracted. The quality of binned assemblies was assessed using QUAST [v4.4; (11)], and taxonomy was assigned to MAGs with the Genome Taxonomy Database (GTDB-Tk v1.7.0) and FastANI (12–17). Finally, MAGs were annotated using the Prokaryotic Genome Annotation Pipeline (PGAP) through National Center for Biotechnology Information [v6.6; (18–20)]. Additional analysis using AntiSMASH v7.0 indicated that WG22 has a complete *mcy* operon that correlates with frequent detections of microcystins in the Winam Gulf, while LS22 is not predicted to produce any common cyanotoxins (21, 22).

AUTHOR AFFILIATIONS

[1]Department of Biological Sciences, Bowling Green State University, Bowling Green, Ohio, USA

[2]Great Lakes Center for Fresh Waters and Human Health, Bowling Green State University, Bowling Green, Ohio, USA

AUTHOR ORCIDs

Katelyn M. Brown http://orcid.org/0000-0002-8581-9490

George S. Bullerjahn 🆔 https://orcid.org/0000-0001-5319-7896

DATA AVAILABILITY

The metagenome-assembled genomes have been deposited in GenBank under the accession numbers JAVSPN000000000 and JAVSPO000000000 for WG22 and LS22, respectively. The versions described in this paper are versions JAVSPN010000000 and JAVSPO010000000. The raw sequence files are available as sequence read archives under SRR25339258 and SRR25339257. The BioProject can be found under the reference PRJNA996591.

REFERENCES

1. Robarts RD, Zohary T. 1987. Temperature effects on photosynthetic capacity, respiration, and growth rates of bloom‐forming cyanobacteria. New Zealand Journal of Marine and Freshwater Research 21:391–399. https://doi.org/10.1080/00288330.1987.9516235
2. Sili C, Torzillo G, Vonshak A. 2012. Arthrospira (Spirulina), p 677–704. In Whitton BA (ed), Ecology of cyanobacteria II: their diversity in space and time. Springer Netherlands, Dordrecht.
3. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. Kbase: the United States department of energy systems biology knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163
4. 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
5. 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
6. 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
7. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165
8. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11:1144–1146. https://doi.org/10.1038/nmeth.3103
9. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843. https://doi.org/10.1038/s41564-018-0171-1
10. 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
11. 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
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
13. Matsen FA, Kodner RB, Armbrust EV. 2010. pplacer: linear time maximum-likelihood and bayesian phylogenetic placement of sequences onto a fixed reference tree. BMC Bioinformatics 11:538. https://doi.org/10.1186/1471-2105-11-538
14. 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
15. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119
16. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490
17. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195
18. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068
19. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz

M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the prokaryotic genome annotation pipeline reach with protein family model curation. Nucleic Acids Res 49:D1020–D1028. https://doi.org/10.1093/nar/gkaa1105

20. 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

21. Blin K, ShawS, AugustijnHE, ReitzZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, HelfrichEJN, van WezelGP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Research 51:W46–W50. https://doi.org/10.1016/j.jviromet.2022.114648

22. Sitoki L, Kurmayer R, Rott E. 2012. Spatial variation of phytoplankton composition, biovolume, and resulting microcystin concentrations in the Nyanza Gulf (Lake Victoria, Kenya). Hydrobiologia 691:109–122. https://doi.org/10.1007/s10750-012-1062-8