Draft Genome Sequence of the Cadmium-Resistant Strain JJU2, Belonging to the Family *Hapalosiphonaceae* of the *Cyanobacteria*

Angelo Joshua Victoria,a Ernelea Cao,a Nico Salmaso,b Nicola Segata,c Claudio Donatib

aInstitute of Biology, University of the Philippines Diliman, Quezon City, Philippines
bResearch and Innovation Centre, Fondazione Edmund Mach, San Michele all'Adige, Italy
cCentre for Integrative Biology, University of Trento, Trento, Italy

ABSTRACT Here, we report the genome of strain JJU2, a cyanobacterium of the family *Hapalosiphonaceae* known to be resistant to high cadmium levels, assembled from a nonaxenic, unialgal culture from Marinduque, Philippines. The draft genome is 7.1 Mb long with a GC content of 40.05% and contains 5,625 protein-coding genes.

The cyanobacterium strain JJU2 is a member of the family *Hapalosiphonaceae* (phylum *Cyanobacteria*) characterized by heterocyte formation and true branching (1). Evidence of cadmium tolerance mechanisms in strain JJU2 were previously reported (2). However, little is known about the molecular genetics of members of the *Hapalosiphonaceae* family, and to date, there are only two genomes sequenced for the genus in public databases, *Hapalosiphon welwitschii* UC-IC-52-3 (3) and *Hapalosiphon* sp. strain MRB220 (4). Differences in secondary metabolite gene clusters were reported by these studies, but the genetic underpinnings of adaptations to metal tolerance are still lacking. To expand the genomic representation of the family *Hapalosiphonaceae*, we report here the genome of strain JJU2, isolated from the heavy-metal-contaminated Mogpog River in Marinduque, Philippines.

Strain JJU2 was isolated from a freely floating colony grown from the isolation of a single filament. The unialgal, nonaxenic culture was maintained in BG-11 medium with a 12-hour light-dark cycle at the Plant Genetics and Cyanobacterial Biotechnology Laboratory of the University of the Philippines, Diliman Institute of Biology. The sample was repeatedly washed with sterile BG-11 medium (5) and filtered through a 0.45-μm membrane (Merck Millipore, USA) prior to extraction using the ZR bacterial/fungal DNA miniprep kit (Zymo, USA). Library preparation and 150-bp paired-end sequencing were performed at the Philippine Genome Center DNA Sequencing Core Facility on the MiSeq platform (Illumina, USA).

Raw sequencing reads were quality trimmed and filtered using FASTX-Toolkit 0.0.13 (6). Contaminant reads were identified by performing k-mer analysis (7) using K-mer Analysis Toolkit 2.3.2 (8) and khmer 2.1.2 (9) (k = 45) and removed using Bowtie 2 2.3.4.1 (10). The 13,891,516 filtered reads were assembled using SPAdes 3.11.0-1 (11) with default parameters and the error-correction pipeline enabled, generating an initial assembly of 1,084 contigs. The genome was predicted by CheckM (12) to be almost complete (96.84%) with virtually no contamination or strain heterogeneity. Scaffolding was performed using MeDuSa (13) with default parameters.

The final assembled 7,145,111-bp-long genome comprised 209 scaffolds with an N_{50} value of 89,291 bp, a GC content of 40.05%, and a coverage of 111×. Annotation and biosynthetic gene cluster prediction was performed with the NCBI Prokaryotic Genome Annotation Pipeline (14) and antiSMASH 4.1.0 (15) using default parameters, identifying

5,625 coding sequences and 41 tRNAs. Genes for cadmium tolerance, such as *czcA*, *czcB*, and *cadA*, were annotated, as were 30 other heavy metal tolerance genes, including the metallothionein *smtA* and its known transcriptional repressor, *smtB*. Prediction of biosynthetic gene clusters revealed that this cyanobacterium has gene clusters for cyanotoxin production with 30% and 27% similarities to microcystin and nostophycin, respectively. It also encodes a polyketide synthase gene cluster with 90% and 71% identities to the ambiguine and welwitindolinone natural-product gene clusters, respectively, highlighting the potential of this species for sensing and adaptation to metal stress and production of toxic metabolites.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number QLKN00000000. The version described in this paper is version QLKN01000000.

REFERENCES

1. Komárek J, Kaštovský J, Mareš J, Johansen JR. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia 86:295–335.
2. de Guzman M, Cao E. 2010. Cadmium binding ability of the blue-green alga Hapalosiphon welwitschii Nägel under controlled conditions. Philipp Sci Lett 3:76–86.
3. Micallef ML, Sharma D, Bunn BM, Gerwick L, Viswanathan R, Moffitt MC. 2014. Comparative analysis of hapalindole, ambiguine and welwitindolinone gene clusters and reconstitution of indole-isonitrile biosynthesis from cyanobacteria. BMC Microbiol 14:213. https://doi.org/10.1186/s12866-014-0213-7.
4. Tan BF, Te SH, Boo CY, Gin KY-H, Thompson JR. 2016. Insights from the draft genome of the subsection V (Stigonematales) cyanobacterium *Hapalosiphon* sp. strain MRB220 associated with 2-MIB production. Stand Genomic Sci 11:58. https://doi.org/10.1186/s40793-016-0175-5.
5. Rippka R. 1988. Isolation and purification of cyanobacteria. Methods Enzymol 167:3–27.
6. Hannon Lab. 2014. FASTX Toolkit. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. http://hannonlab.cshl.edu/fastx_toolkit/.
7. Zhang Q, Pell J, Canino-Koning R, Howe AC, Brown CT. 2014. These are not the k-mers you are looking for: efficient online k-mer counting using a probabilistic data structure. PLoS One 9:e101271. https://doi.org/10.1371/journal.pone.0101271.
8. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a k-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics 33:574–576. https://doi.org/10.1093/bioinformatics/btw663.
9. Crusoe MR, Alameldin HF, Awad S, Boucher E, Caldwell A, Cartwright R, Charbonneau A, Constantinides B, Edvenson G, Fay S, Fenton J, Fenzl T, Fish J, Garcia-Gutierrez L, Garland P, Gluck J, González I, Guermond S, Guo J, Gupta A, Herr JR, Howe A, Hyer A, Härpfer A, Irber L, Kidd R, Lin D, Lippi J, Mansour T, McA'Nulty P, McDonald E, Mizzi J, Murray KD, Nahum JR, Nanlohy K, Nederbragt AJ, Ortiz-Zuazaga H, Ory J, Pell J, Pepe-Ranney C, Russ ZN, Schwarz E, Scott C, Seaman J, Sievert S, Simpson J, Skennerton CT, Spencer J, Srinivasan R, Standage D, et al. 2015. The khmer software package: enabling efficient nucleotide sequence analysis [version 1; referees: 2 approved, 1 approved with reservations]. F1000Research 4:900. https://doi.org/10.12688/f1000research.6924.1.
10. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
12. 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.
13. Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, Lió P, Crescenzi P, Fani R, Fondi M. 2015. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31:2443–2451. https://doi.org/10.1093/bioinformatics/btv171.
14. 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.
15. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res 45:W36–W41. https://doi.org/10.1093/nar/gkx319.