Draft Genome Sequence of *Okeania* sp. Strain KiyG1, Assembled from Single-Amplified Genomes Collected from Cape Kiyan, Okinawa, Japan

Muhammad Wahyudin Lewaru, Yohei Nishikawa, Keigo Ide, Masato Kogawa, Masahito Hosokawa, Ashok Zachariah Samuel, Shinpei Sumimoto, Handung Nuryadi, Shoichiro Suda, Haruko Takeyama

*Department of Life Science and Medical Bioscience, Waseda University, Tokyo, Japan
Computational Bio Big-Data Open Innovation Laboratory, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan
Research Organization for Nano and Life Innovation, Waseda University, Tokyo, Japan
Faculty of Science, University of the Ryukyus, Okinawa, Japan
Institute for Advanced Research of Biosystem Dynamics, Waseda Research Institute for Science and Engineering, Tokyo, Japan

Muhammad Wahyudin Lewaru and Yohei Nishikawa contributed equally to this work as co-first authors. Author order was determined on the basis of increasing seniority.

**ABSTRACT** The genus *Okeania* is a globally distributed group of microorganisms that live in shallow seabed regions. These organisms play several environmentally important roles and are also known producers of several active secondary metabolites with potential human applications. Here, we present a draft genome of *Okeania* sp. strain KiyG1 (92.7% completeness) that was assembled from four single-amplified genomes.

Cyanobacteria play several important ecological roles in marine ecosystems, particularly contributing to nitrogen fixation and reduction of global carbon flux (1). They are also known producers of several biologically active secondary metabolites (2). However, their genomic details remain incomplete. The filamentous cyanobacteria, including *Okeania* spp. (defined in 2013), usually form biological mats composed of various kinds of bacteria (3), often with relatively lower abundance of the target cyanobacteria, which causes DNA contamination and reduces the accuracy of genome sequencing (4, 5). Therefore, determining the genome sequence from isolated single cells is advantageous (6, 7).

Cyanobacteria of *Okeania* sp. strain KiyG1 (GenBank accession number LC534992) were collected from the shallow seabed (water depth of 10 cm) of Cape Kiyan in Okinawa, Japan (26°04’52.6”N, 127°39’25.5”E), on 15 June 2019. The samples were washed with artificial seawater, cut into small pieces with a scalpel, and then treated with 5% sodium dodecyl sulfate (Sigma-Aldrich, St. Louis, MO, USA) in Dulbecco’s phosphate-buffered saline (DPBS) (Thermo Fisher Scientific, Waltham, MA, USA) for about 5 min at room temperature. This mixture was then filtered through a membrane filter (MF-Millipore filter with 5-μm pore size, SMWP04700; Millipore, Merck, Germany), and the filamentous cyanobacteria trapped on the membrane were transferred into 1 ml DPBS. The mixture was then vortex-mixed to disentangle the cyanobacteria to single cells. Eight single cells were picked from the vortex-mixed sample by pipetting (Drummond Scientific, Broomall, PA, USA) during observation through a microscope (CKX53, with magnification of ×10; Olympus, Tokyo, Japan). The cells were then transferred to a 384-well plate containing 0.8 μl DPBS and were physically ground using a micropipette tip (123R-254CS; Fukae Kasei Co., Japan). The whole genome in the resulting samples was amplified with the REPLI-g single-cell kit (Qiagen Inc., Valencia, CA, USA) at 30°C for 120 min. Sufficient DNA yield (>1 ng) for sequencing was available for four samples. The independent sequencing libraries were prepared from each of the four samples with the Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA).
USA) according to the manufacturer’s instructions and were sequenced on an Illumina MiSeq system for 75 cycles of paired-end sequencing, generating 116 to 145 Mbp. 

Low-quality reads (>50% of bases with quality scores of <25) were eliminated, and the 3’ ends (quality scores of <20) were trimmed using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit) and PRINSEQ (8). The cleaned reads (>20 bp) were assembled using SPAdes v3.13.1 (9) with the following settings: careful, sc. We successfully acquired four single-cell genomes. Because the sequences of >100 single-copy marker genes identified with CheckM v1.0.18 (10) exhibited >99.99% identity, DNA from the four amplified genomes was mixed evenly and sequenced with a GridION system (Oxford Nanopore Technologies Ltd., Oxford, UK) using the rapid sequencing kit (SQK-RAD004; Oxford Nanopore Technologies Ltd.), generating 1.13 Gbp. The long-read draft genome was assembled using Flye v2.7.1 (11) and polished with Pilon v1.22 (12).

The polished long-read draft genome of Okeania sp. strain KiyG1 was evaluated with QUAST v5.0.2 (13) and CheckM v1.0.18 and annotated with Prokka v1.13 (14). The genome completeness was 92.74%, and the contamination was 2.91%. Additional genome statistics for KiyG1 are provided in Table 1. Default parameters were used for all software unless otherwise noted.

**Data availability.** The assembled genomes reported in this article were deposited in DDBJ under the accession number SAMD00236478. The raw reads are also available, under the accession numbers SAMD00233784 to SAMD00233787 (for Illumina reads) and SAMD00233788 (for Nanopore reads).

**ACKNOWLEDGMENTS**

We acknowledge financial support from a JST CREST grant (grant JPMJCR12A4) and a Grant-in-Aid for Scientific Research (S) (grant 17H06158).

The supercomputing resources were provided by the Human Genome Center (University of Tokyo).

**REFERENCES**

1. Charpy L, Casareto BE, Langlade MJ, Suzuki Y. 2012. Cyanobacteria in coral reef ecosystems: a review. Rev J Marine Biol 2012:259571. https://doi.org/10.1155/2012/259571.

2. Nunnery JK, Mevers E, Gerwick WH. 2010. Biologically active secondary metabolites from marine cyanobacteria. Curr Opin Biotechnol 21:787–793. https://doi.org/10.1016/j.copbio.2010.09.019.

3. Prieto-Barajas CM, Valencia-Cantero E, Santoyo G. 2018. Microbial mat ecosystems: structure types, functional diversity, and biotechnological application. Electron J Biotechnol 31:48–56. https://doi.org/10.1016/j.ejbt.2017.11.001.

4. Engene N, Paul VJ, Byrum T, Gerwick WH, Thor A, Ellisman MH. 2013. Five chemically rich species of tropical marine cyanobacteria of the genus Okeania gen. nov. (Oscillatoriales, Cyanoprokaryaota). J Phycol 49:1095–1106. https://doi.org/10.1111/jpy.12115.

5. Moss NA, Leao T, Rankin MR, McCullough TM, Qu P, Korobeinikov A, Smith JL, Gerwick L, Gerwick WH. 2018. Ketoreductase domain dysfunction expands chemodiversity: malyngamide biosynthesis in the cyanobacterium Okeania hirsuta. ACS Chem Biol 13:3385–3395. https://doi.org/10.1021/acschembio.8b00910.

6. Hosokawa M, Nishikawa Y, Kogawa M, Takeyama H. 2017. Massively parallel whole genome amplification for single-cell sequencing using droplet microfluidics. Sci Rep 7:5199. https://doi.org/10.1038/s41598-017-05436-4.

7. Chijiwa R, Hosokawa M, Kogawa M, Nishikawa Y, Ide K, Sakashib C, Takahashi K, Takeyama H. 2020. Single-cell genomics of uncultured bacteria reveals dietary fiber responders in the mouse gut microbiota. Microbiome 8:5. https://doi.org/10.1186/s40168-019-0779-2.

8. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.

**TABLE 1** Assembly and quality statistics for the Okeania sp. strain KiyG1 single-amplified genome

| Assembly statistic      | Result for Okeania sp. strain KiyG1 |
|-------------------------|--------------------------------------|
| No. of contigs          | 198                                  |
| Total contig length (bp) | 10,982,302                           |
| N50 (bp)                | 79,819                               |
| GC content (%)          | 36.81                                |
| Completeness (%)        | 92.74                                |
| Contamination (%)       | 2.91                                 |
| Predicted genome size (bp) | 11,842,034                          |
| No. of coding sequences | 9,096                                |
| No. of tRNA genes       | 33                                   |

a Estimated with QUAST.

b Estimated with CheckM.

c Estimated with Prokka.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pyrjhelski 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.

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. 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.

12. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.

13. 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.

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