



Draft Genome Sequence of the Cyanobacterium *Synechococcus* sp. Strain Nb3U1

Christopher L. Pierpont,^a Satoshi Ohkubo,^b Hideaki Miyashita,^c Scott R. Miller^a

^aDivision of Biological Sciences, University of Montana, Missoula, Montana, USA

^bGraduate School of Life Sciences, Tohoku University, Sendai, Aoba-ku, Japan

^cGraduate School of Human and Environmental Studies, Kyoto University, Kyoto, Sakyo-ku, Japan

ABSTRACT We report the 3.5-Mb draft genome sequence of the cyanobacterium *Synechococcus* sp. strain Nb3U1, which was isolated from a microbial mat sample collected from Nakabusa Hot Spring, Nagano, Japan.

Synechococcus sp. strain Nb3U1 was isolated from a microbial mat sample collected from Nakabusa Hot Spring, Nagano, Japan; mat homogenates were grown in BG11 medium under far-red light, and individual cyanobacterial cells were isolated microscopically from these enrichments (1). 16S rRNA sequence analysis identified Nb3U1 as a member of lineage T1 (2). T1 cyanobacteria are a cosmopolitan but poorly understood group from mesothermic to moderately high-temperature habitats in alkaline geothermal environments. Notably, T1 cyanobacteria appear to be the less-thermotolerant sister taxa of the *Synechococcus* A/B clade, members of which define the upper temperature limit for phototrophy on Earth (~74°C [3]).

Strain Nb3U1 was grown in liquid BG11 medium under ~75 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ of cool white fluorescent light with a 12-h/12-h photoperiod. Genomic DNA (gDNA) for paired-end, short-read sequencing was extracted from the growing cells using Qiagen's DNeasy PowerBiofilm kit. A library was prepared using a Nextera DNA Flex kit, followed by 150 cycles of paired-end sequencing on an Illumina NextSeq 550 platform. High-molecular-weight gDNA for long-read sequencing was later extracted from the same culture using Qiagen's Genomic-tip 20/G protocol, prepared using the Nanopore ligation sequencing kit without shearing or size selection and sequenced for 48 h using a Nanopore MinION sequencer (FLO-MIN106D, R9.4.1); GUPPY v4.5.4 (<https://staff.aist.go.jp/yutaka.ueno/guppy/>) was used for Nanopore base calling. Prior to sequencing, DNA quality and quantity were assessed by spectrophotometry (Agilent Tape Station) and fluorimetry (Qubit 2.0), respectively. The quality of all sequenced reads was checked using FastQC v0.11 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), but none were removed. A hybrid draft assembly using 8,011,914 paired-end Illumina short reads (35 to 151 bp) and 141,038 Nanopore long reads (83 to 51,308 bp) was built *de novo* using SPAdes v3.12.0 (4). Following the assembly, contigs shorter than 1 kbp were removed prior to subsequent analyses. Kraken v2.1.2 (5) and BLAST+ v2.2.31 (6) were used to identify and extract contigs with high sequence similarity to cyanobacteria if (i) the contig returned significant local BLAST hits (E values, $\leq 1e^{-100}$) to two *Synechococcus* A/B reference genomes (7) and/or (ii) Kraken assigned an NCBI TaxID from the phylum *Cyanobacteria*, the order *Synechococcales*, or either of the above reference genomes.

Genomic statistics for these extracted contigs were measured using QUAST v4.5 (8). The total size of the refined assembly is 3,487,976 bp in 4 contigs, with an N_{50} value of 2,215,988 bp, a GC content of 55.19%, and a mean coverage of 73.0 \times . The Nb3U1 genome was estimated to be 93.8% complete using BUSCO v5.2.2 (9) with reference to the *Synechococcales* lineage. In identical analyses, the closed *Synechococcus* A/B reference genomes were estimated to be 94.4% and 94.9% complete. We therefore conclude that the

Editor Julia A. Maresca, University of Delaware

Copyright © 2022 Pierpont et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Scott R. Miller, scott.miller@umontana.edu.

The authors declare no conflict of interest.

Received 10 March 2022

Accepted 8 April 2022

Published 19 April 2022

Nb3U1 genome assembly is essentially complete. The refined assembly was annotated using NCBI's PGAP software (10). Of the 3,300 predicted features, 3,209 are protein coding, 42 are tRNAs, 4 are noncoding RNAs (ncRNAs), and 3 are rRNAs. All software tools were run with default parameters unless otherwise specified.

Data availability. This whole-genome shotgun sequencing project has been deposited at DDBJ/ENA/GenBank under accession number [JAKFYQ000000000.1](https://doi.org/10.1093/ismej/2017.98). The version described in this paper is version [JAKFYQ010000000](https://doi.org/10.1007/s00792-008-0205-3). These data are associated with BioProject accession number [PRJNA795194](https://doi.org/10.1093/ismej/2017.98) under BioSample accession number [SAMN24695622](https://doi.org/10.1093/ismej/2017.98). The raw data from the Illumina and Nanopore sequencing are available under SRA accession numbers [SRR17483484](https://doi.org/10.1093/ismej/2017.98) and [SRR17483483](https://doi.org/10.1093/ismej/2017.98), respectively.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation award IOS-1755407 to S.R.M.

We thank Tim Wheeler for Nanopore library preparation and sequencing, and two anonymous reviewers for their helpful comments.

We declare no conflicts of interest.

REFERENCES

- Ohkubo S, Miyashita H. 2017. A niche for cyanobacteria producing chlorophyll *f* within a microbial mat. *ISME J* 11:2368–2378. <https://doi.org/10.1038/ismej.2017.98>.
- Lau MCY, Aitchison JC, Pointing SB. 2009. Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles* 13:139–149. <https://doi.org/10.1007/s00792-008-0205-3>.
- Ward DM, Castenholz RW, Miller SR. 2012. Cyanobacteria in geothermal habitats, p 39–63. *In* Whitton BA (ed), *Ecology of cyanobacteria II: their diversity in space and time*, 1st ed. Springer, Dordrecht, Netherlands.
- Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes de novo assembler. *Curr Protoc Bioinf* 70:e102. <https://doi.org/10.1002/cpbi.102>.
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
- Bhaya D, Grossman AR, Steunou A, Khuri N, Cohan FM, Hamamura N, Melendrez MC, Bateson MM, Ward DM, Heidelberg JF. 2007. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J* 1:703–713. <https://doi.org/10.1038/ismej.2007.46>.
- 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>.
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>.
- 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* 49:1020–1028. <https://doi.org/10.1093/nar/gkaa1105>.