



Complete Genome Sequence of the Edible Filamentous Cyanobacterium *Arthrospira platensis* NIES-39, Based on Long-Read Sequencing

 Hideaki Shiraishi,^a Haruka Nishida^a

^aGraduate School of Biostudies, Kyoto University, Kyoto, Japan

ABSTRACT *Arthrospira platensis* is a filamentous cyanobacterium that is cultivated and used worldwide as a source of food and food additives. Here, we report the complete genome sequence (6,818,916 bp) of *A. platensis* NIES-39, one of the model strains of *A. platensis*, providing an improved reference genome sequence for this strain.

A *Arthrospira platensis* is an edible cyanobacterium that is also known as spirulina (1, 2). NIES-39 is one of the model strains, and its draft genome sequence has been reported (3). However, the genome consists of 18 contigs separated by gaps. To better understand this strain's genetic properties, we decided to determine its full-length genome sequence.

The strain was obtained in 2008 from the Microbial Culture Collection at the National Institute for Environmental Studies (MCC-NIES) (Tsukuba, Japan), where it was maintained as a live culture. We also maintained it as a live culture until 2016. Then, it was cryopreserved since the conditions for cryopreservation were found (4). Cells were propagated from a single trichome in modified SOT medium (5, 6) and ground in liquid nitrogen using a mortar and pestle. DNA was extracted with phenol and purified twice by CsCl density gradient centrifugation (7, 8). The DNA was further purified using 0.87 volume of ProNex Size-Selective Chemistry reagent (Promega). The purified DNA was sonicated in a g-TUBE (Covaris) to obtain 10- to 20-kb fragments. A library was prepared using the SMRTbell Express template preparation kit v.2.0 (Pacific Biosciences [PacBio]). Sequence data were obtained using the Sequel IIe system (PacBio) and imported into SMRT Link v.10.1.0.119528 to obtain high-fidelity (HiFi) reads with quality values of >20 or 99% accuracy, which yielded 33,140 reads. Using Filtlong v.0.2.0 (<https://github.com/rwwick/Filtlong>), HiFi reads of >1,000 bp were selected, and the worst 10% of reads were discarded (parameters: --min_length 1000 --keep_percent 90). The resultant 29,202 reads (average length, 7,669 bp; N_{50} , 8,626 bp) were used to assemble contigs with Flye v.2.9 (9), which yielded three contigs (3,451 kb, 2,846 kb, and 549 kb in length, with 33-fold coverage). Default parameters were used for all software unless otherwise noted.

All contigs contained members of dispersed repetitive sequences at both ends. To examine the relationships between the contigs, we synthesized six PCR primers that bind unique sequences near the termini of the contigs (Table 1 and

TABLE 1 List of PCR primers

Primer	Nucleotide sequence	Binding site ^a
Primer 1	5'-GCTTTGGGAGTGTTGGGAAAATTTTTGA-3'	2482240–2482211
Primer 2	5'-ATGATGTTGGGGATGTTTAGATACTTTG-3'	5904111–5904140
Primer 3	5'-ACTCCTAATAAACCTAACTTTCTCCT-3'	5923425–5923396
Primer 4	5'-GCCAAGTGTTAAGGCAGAGGATTGTGGT-3'	1922043–1922072
Primer 5	5'-ACCACCATGATCATAACCATGATCATAGCC-3'	1945954–1945925
Primer 6	5'-GCCATCCAAGAATTTTAGACGGTCTGGT-3'	2461026–2461055

^a Nucleotide positions in the finally determined full-length genome sequence are shown.

Editor Julia A. Maresca, University of Delaware

Copyright © 2022 Shiraishi and Nishida. 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 Hideaki Shiraishi, shiraishi.hideaki.8e@kyoto-u.ac.jp.

The authors declare no conflict of interest.

Received 27 October 2022

Accepted 1 December 2022

Published 20 December 2022

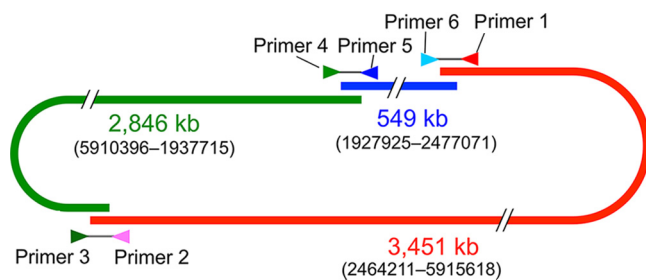


FIG 1 Schematic representation of junctional PCR. Numbers in parentheses show nucleotide positions in the finally determined full-length sequence.

Fig. 1). PCR amplification of the genomic DNA using KOD One PCR Master Mix (Toyobo) resulted in efficient amplification of contig junctions when three specific primer pairs were used, as summarized in Fig. 1. Analysis of the PCR-amplified DNAs with restriction enzymes indicated that, in the genomic DNA, each terminus of the contigs was connected to an adjacent one sharing the terminal sequence. To determine the precise nucleotide sequences of the contig junctions, we searched the FASTQ file for the HiFi reads containing the specific unique primer sequences used in the PCR mentioned above (Table 1). Through manual assembly of the extracted HiFi reads corresponding to each junction, the nucleotide sequences of the three contig junctions were unambiguously determined.

The genome of *A. platensis* NIES-39 was a circular DNA of 6,818,916 bp, with a GC content of 44.3%. Genome annotation using DFAST v.1.2.18 (10) with CyanoBase (11) as an optional database predicted 6,373 protein-coding sequences, 4 rRNA genes, and 48 tRNA genes. Among the newly found genes were the porphobilinogen deaminase gene (*hemC*), which is involved in heme biosynthesis, and a cluster of gas vesicle genes (*gvpK-gvpL/gvpF-gvpG*).

Data availability. The complete genome sequence has been deposited in GenBank under the accession number [AP026945](https://doi.org/10.1093/bioinformatics/btx713). The sequencing data can be found in the DDBJ Sequence Read Archive under the accession number [DRR411675](https://doi.org/10.1093/bioinformatics/btx713).

REFERENCES

- Belay A. 2013. Biology and industrial production of *Arthrospira* (*Spirulina*), p 339–358. In Richmond A, Hu Q (ed), *Handbook of microalgal culture*, 2nd ed. Wiley-Blackwell, Hoboken, NJ.
- Furmaniak MA, Misztak AE, Franczuk MD, Wilmotte A, Waleron M, Waleron KF. 2017. Edible cyanobacterial genus *Arthrospira*: actual state of the art in cultivation methods, genetics, and application in medicine. *Front Microbiol* 8: 2541. <https://doi.org/10.3389/fmicb.2017.02541>.
- Fujisawa T, Narikawa R, Okamoto S, Ehira S, Yoshimura H, Suzuki I, Masuda T, Mochimaru M, Takaichi S, Awai K, Sekine M, Horikawa H, Yashiro I, Omata S, Takarada H, Katano Y, Kosugi H, Tanikawa S, Ohmori K, Sato N, Ikeuchi M, Fujita N, Ohmori M. 2010. Genomic structure of an economically important cyanobacterium, *Arthrospira* (*Spirulina*) *platensis* NIES-39. *DNA Res* 17:85–103. <https://doi.org/10.1093/dnares/dsq004>.
- Shiraishi H. 2016. Cryopreservation of the edible alkalophilic cyanobacterium *Arthrospira platensis*. *Biosci Biotechnol Biochem* 80:2051–2057. <https://doi.org/10.1080/09168451.2016.1189320>.
- Shiraishi H, Tabuse Y. 2013. The *AplI* restriction-modification system in an edible cyanobacterium, *Arthrospira* (*Spirulina*) *platensis* NIES-39, recognizes the nucleotide sequence 5′-CTGCAG-3′. *Biosci Biotechnol Biochem* 77:782–788. <https://doi.org/10.1271/bbb.120919>.
- Tadama S, Shiraishi H. 2017. Growth of the edible microalga *Arthrospira platensis* in relation to boron supply. *Int J GEOMATE* 12:90–95. <https://doi.org/10.21660/2017.30.2580>.
- Radloff R, Bauer W, Vinograd J. 1967. A dye-buoyant-density method for the detection and isolation of closed circular duplex DNA: the closed circular DNA in HeLa cells. *Proc Natl Acad Sci U S A* 57:1514–1521. <https://doi.org/10.1073/pnas.57.5.1514>.
- Maniatis T, Fritsch EF, Sambrook J. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. <https://doi.org/10.1002/abio.370050118>.
- 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>.
- Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 34: 1037–1039. <https://doi.org/10.1093/bioinformatics/btx713>.
- Fujisawa T, Narikawa R, Maeda S, Watanabe S, Kanesaki Y, Kobayashi K, Nomata J, Hanaoka M, Watanabe M, Ehira S, Suzuki E, Awai K, Nakamura Y. 2017. CyanoBase: a large-scale update on its 20th anniversary. *Nucleic Acids Res* 45:D551–D554. <https://doi.org/10.1093/nar/gkw1131>.