



Draft Genome Sequence of the Green Microalga *Chlorella* sp. Strain BAC9706, Isolated from Lake Baikal, Russia

 Ivan S. Petrushin,^{a,b} Sergei I. Belikov,^a Olga I. Belykh,^a Irina Tikhonova,^a Lubov I. Chernogor^a

^aLimnological Institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia

^bIrkutsk State University, Irkutsk, Russia

ABSTRACT Green algae of the phylum *Chlorophyta* are the most widespread autotrophic picoplankton in Lake Baikal (Russia). To expand our molecular biological knowledge of these microalgae and compare them in the future with an endosymbiotic strain, we present here the draft genome sequence of *Chlorella* sp. strain BAC9706.

Green algae of the phylum *Chlorophyta* are the most widespread autotrophic picoplankton in Lake Baikal (Russia) (1–3). The related algae are symbionts of freshwater sponges and were assigned to the genus *Choricystis* based on analysis of the 18S rRNA and *rbcL* genes (4, 5). To expand our molecular biological knowledge of these microalgae and compare them in the future with an endosymbiotic strain, we present a draft version of the genome sequence of the Baikal autotrophic picoplankton green alga *Chlorella* sp. strain BAC9706.

Water samples containing microalgae were collected from the pelagic zone in the southern part of Lake Baikal, 5 km from Listvyanka (51.47°38' N, 104.55°54' E). The pure strain was obtained by serial transfers on BG-11 medium-enriched agar plates. Individual colonies of algae were cultivated in f/2 medium in 0.5-liter spin flasks, under constant illumination (200 $\mu\text{E m}^{-2} \text{s}^{-1}$ white fluorescent light). The dense culture was harvested by centrifugation (300 $\times g$, 15 min); the pellet was rapidly transferred and stored at -80°C until DNA isolation. In this study, genomic DNA of the green microalga *Chlorella* sp. strain BAC9706 was extracted from microalgal cultures using the Xpert directXtract plant lysis kit (GRiSP, Portugal).

Whole-genome sequencing was performed using the MiSeq platform with paired-end chemistry (2 \times 250 bp) (Illumina, USA). A DNA library was prepared from genomic DNA using a Nextera DNA Flex library prep kit (Illumina). A total of 12,987,796 paired-end reads was obtained, giving a coverage depth of 60 \times . The hypervariable region of the 18S rRNA gene sequence was assembled with GetOrganelle v. 1.6.2e (6) (with default settings). Strain BAC9706 showed the highest 18S rRNA gene phylogenetic affiliation to species of the phylum *Chlorophyta* belonging to the genus *Chlorella*.

A draft assembly was built using SPAdes v. 3.13.0 with default settings, raw read filtering, and error correction with a built-in BayesHammer module (quality threshold, 98%) (7). This draft assembly contained 5,837 contigs with an N_{50} value of 44,654 bp, and the largest contig was 317,606 bp long.

The resulting contigs were combined into scaffolds using Ragout v. 2.3 with default settings (<https://github.com/fenderglass/Ragout>) (8) using *Chlorella vulgaris* strain UTEX259 (GenBank accession no. [VATW01000000](https://www.ncbi.nlm.nih.gov/nuccore/VATW01000000)) and *Chlorella vulgaris* strain UTEX 395 (GenBank accession no. [LDKB01000000](https://www.ncbi.nlm.nih.gov/nuccore/LDKB01000000)) as the reference genomes. We selected these two genomes from all the whole-genome sequences of *Chlorella* strains available in GenBank because the highest percentages of raw reads were mapped onto these genomes using Bowtie.

The final genome assembly had a total length of 32,957,766 bp, resulting in 201

Citation Petrushin IS, Belikov SI, Belykh OI, Tikhonova I, Chernogor LI. 2020. Draft genome sequence of the green microalga *Chlorella* sp. strain BAC9706, isolated from Lake Baikal, Russia. *Microbiol Resour Announc* 9:e00966-20. <https://doi.org/10.1128/MRA.00966-20>.

Editor Jason E. Stajich, University of California, Riverside

Copyright © 2020 Petrushin 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 Ivan S. Petrushin, ivan.kiel@gmail.com.

Received 17 August 2020

Accepted 2 October 2020

Published 22 October 2020

scaffolds (N_{50} , 521,038 bp; L_{50} , 20), with a GC content of 61.92%; the largest scaffold was 1,817,640 bp.

Genome completeness analysis with BUSCO v. 3.1.0 using the “chlorophyta_odb10” data set with 2,168 benchmarking universal single-copy orthologs (BUSCOs) and default settings (9) showed results of 79.2% complete, 4.8% fragmented, and 16% missing. For the genome sequences of UTEX259 and UTEX 395, used as references, BUSCO reported 88.4% and 79.3% complete BUSCOs, respectively.

The presented draft nuclear genome assembly of *Chlorella* sp. strain BAC9706 provides a strong basis for comparative genetic analyses and will help elucidate metabolic processes.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession no. [JABWAE000000000](https://www.ncbi.nlm.nih.gov/nuccore/JABWAE000000000). The raw reads are available via BioProject no. [PRJNA591660](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA591660) and SRA no. [SRX7258937](https://www.ncbi.nlm.nih.gov/sra/SRX7258937).

ACKNOWLEDGMENTS

This material is based upon work supported by the Russian Foundation for Basic Research (RFBR) under grant no. 18-04-00224. Sample collection was carried out within the framework of basic budget funding no. 0345-2019-0002 (AAAA-A16-116122110066-1). The bioinformatics data analysis was performed in part on the equipment of the Bioinformatics Shared Access Center at the Federal Research Center Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (ICG SB RAS).

REFERENCES

- Nagata T, Takai K, Kawanobe K, Kim D-S, Nakazato R, Guselnikova N, Bondarenko N, Mologawaya O, Kostrnova T, Drucker V, Satoh Y, Watanabe Y. 1994. Autotrophic picoplankton in southern Lake Baikal: abundance, growth and grazing mortality during summer. *J Plankton Res* 16:945–959. <https://doi.org/10.1093/plankt/16.8.945>.
- Belykh OI, Semenova EA, Kuznedelov KD, Zaika EI, Guselnikova NE. 2000. A eukaryotic alga from picoplankton of Lake Baikal: morphology, ultrastructure and rDNA sequence data. *Hydrobiologia* 435:83–90. <https://doi.org/10.1023/A:1004056604534>.
- Fietz S, Bleiß W, Hepperle D, Koppitz H, Krienitz L, Nicklisch A. 2005. First record of *Nannochloropsis limnetica* (Eustigmatophyceae) in the autotrophic picoplankton from Lake baikal. *J Phycol* 41:780–790. <https://doi.org/10.1111/j.0022-3646.2005.04198.x>.
- Belykh OI, Ekaterina G, Sorokovikova T, Saphonova A, Tikhonova IV. 2006. Autotrophic picoplankton of Lake Baikal: composition, abundance and structure. *Hydrobiologia* 568:9–17. <https://doi.org/10.1007/s10750-006-0340-8>.
- Kulakova NV, Kashin SA, Bukin YS. 2020. The genetic diversity and phylogeny of green microalgae in the genus *Choricystis* (Trebouxiophyceae, Chlorophyta) in Lake Baikal. *Limnology (Tokyo)* 21:15–24. <https://doi.org/10.1007/s10201-019-00587-x>.
- Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* 21:241. <https://doi.org/10.1186/s13059-020-02154-5>.
- Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Pribelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, McLean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling single-cell genomes and minimetagenomes from chimeric MDA products. *J Comput Biol* 20:714–737. <https://doi.org/10.1089/cmb.2013.0084>.
- Kolmogorov M, Armstrong J, Raney BJ, Streeter I, Dunn M, Yang F, Odom D, Flicek P, Keane TM, Thybert D, Paten B, Pham S. 2018. Chromosome assembly of large and complex genomes using multiple references. *Genome Res* 28:1720–1732. <https://doi.org/10.1101/gr.236273.118>.
- Waterhouse RM, Seppely M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35:543–548. <https://doi.org/10.1093/molbev/msx319>.