

UC Berkeley

UC Berkeley Previously Published Works

Title

Genome assembly and annotation of microalga *Nannochloropsis oceanica* C018

Permalink

<https://escholarship.org/uc/item/5w41x9kr>

Journal

Microbiology Resource Announcements, 14(2)

ISSN

2169-8287

Authors

Estrada-Graf, Adrian

Koneru, Hari

Arnold, Jason

et al.

Publication Date

2025-02-11

DOI

10.1128/mra.00883-24

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Genome assembly and annotation of microalga *Nannochloropsis oceanica* C018

Adrian Estrada-Graf,¹ Hari Koneru,¹ Jason Arnold,² Sara Calhoun,³ Igor V. Grigoriev,^{3,4} Zackary I. Johnson^{1,5}

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT The microalga *Nannochloropsis* is an important organism for algae-based biocommodity production of food, feed, and fuel, among other products. Using PacBio Revio, we sequenced, assembled, and annotated a 26.41 Mbp *Nannochloropsis oceanica* C018 genome.

KEYWORDS algae, carbon capture, biofuels, next-generation sequencing, biotechnology, bioinformatics

Microalgae-based technologies have emerged as sustainable approaches to help mitigate the impact of rising CO₂ emissions while providing scalable solutions applicable in agriculture, medicine, and energy, among others. *Nannochloropsis* microalgae exhibit high productivity and tolerance for elevated CO₂ concentrations, thrive in outdoor conditions, have potential for biofuel production, and have emerged as model organisms for applied phycology (1, 2). Additionally, the biochemical characteristics of *Nannochloropsis* make them a valuable candidate as a food source for both animals and humans. *Nannochloropsis oceanica* C018 was isolated by S Brown (University of Hawaii-Manoa) near the Island of Hawaii (19.666° -156.032°) using dilution to extinction with seawater-based F/2 medium (3).

Nannochloropsis oceanica C018 was grown in 20 L carboys with F/2 (3) on a 12-hour day/night light cycle (~400 μmol Q/m²/sec) at 25°C and bubbled with sterile air. Following cell harvesting, DNA was isolated with Quick-DNA/RNA Miniprep Kit D7001 (Zymo Research, USA). Total DNA was quantified using DeNovix DS-11 Series Spectrophotometer/Fluorometer (DeNovix Inc., USA) and normalized to 45 ng/μL prior to sequencing library preparation. DNA fragments between 15 and 20 kb were selected for sequencing, and libraries were prepared using Express Template Prep kit version 3 (Pacific Biosciences, USA) for standard HiFi library preparation. High-fidelity sequencing reads were obtained with Pacific Biosciences Revio system with version SMRT Link v13 and Instrument Control Software (ICS) v13. 1,906,268 total reads were generated with an N50 value of 5090 bp, and HiFi read yield was 11,001,781,445 bp.

Genome assembly followed a modified version of the VGP pipeline (4) in usegalaxy.org (5). Samtools v.1.15.1 (6) was used to extract fastq files from .bam sequence output files, and Cutadapt v.4.6 (7) was implemented to trim adapter sequences and filter low-quality reads. Hifiasm v.0.19.8 (8) was used to assemble contigs from high-fidelity reads with “light” purge level. The resulting assembly was decontaminated with kraken2 v.2.1.1 (9), and duplicates were purged with purge_dups v.1.2.6 (10). Completeness of the resulting genome was estimated with BUSCO v.5.5.0 (11) and quality assessed with QUAST v.5.2.0 (12). Default parameters were used except where otherwise noted.

The resulting genome assembly consisted of 30 contigs, 26.41 Mbp (415× coverage) with a GC content of 54.26%, and a L₅₀ of 10. The genome was annotated following the JGI Annotation pipeline and PhycoCosm (12), with 80 million paired-end transcriptomic reads. Functional annotation of the 8,768 gene models was

Editor Jason E. Stajich, University of California
Riverside, Riverside, California, USA

Address correspondence to Zackary I. Johnson,
zij@duke.edu.

The authors declare no conflict of interest.

Received 28 August 2024

Accepted 9 December 2024

Published 27 December 2024

Copyright © 2024 Estrada-Graf et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

performed using the JGI annotation pipeline. Both genome assembly size and gene count are within the range of other published *Nannochloropsis* genomes (13–15) (<https://phycocosm.jgi.doe.gov/eustigmatophyta>). This genome will spur the continued development of microalgae for use as a biofuel feedstock and provide prerequisite information needed for genetic manipulation and expression analyses.

ACKNOWLEDGMENTS

This work was supported by U.S. Department of Energy grant numbers DE-EE0007091, DE-EE0008518, and DE-EE0009278 and U.S. Department of Agriculture (through Cornell University) grant number 2019-69012-29905 to ZIJ and benefited from the entire “MAGIC” team. This project was also funded by the Duke Bass Connections program. AEG was supported by a U.S./Mexico Fulbright García-Robles award. The work (proposal: 10.46936/10.25585/60008191) conducted by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. Duke Sequencing and Genomic Technologies would like to acknowledge funding from NIH S10OD034222-01 for data obtained with Revo system.

AUTHOR AFFILIATIONS

¹Marine Laboratory, Duke University, Beaufort, North Carolina, USA

²Department of Molecular Genetics and Microbiology, Duke Microbiome Center, Duke University, Durham, North Carolina, USA

³US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA

⁴Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, California, USA

⁵Biology, Civil & Environmental Engineering and Duke Microbiome Center, Duke University, Durham, North Carolina, USA

AUTHOR ORCIDs

Adrian Estrada-Graf  <http://orcid.org/0000-0002-4862-4176>

Zackary I. Johnson  <http://orcid.org/0000-0003-0793-8512>

AUTHOR CONTRIBUTIONS

Adrian Estrada-Graf, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Writing – original draft | Hari Koneru, Investigation, Writing – review and editing | Jason Arnold, Conceptualization, Methodology, Supervision, Writing – review and editing | Sara Calhoun, Formal analysis, Resources | Igor V. Grigoriev, Funding acquisition, Writing – review and editing | Zackary I. Johnson, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

This whole genome project has been deposited at GenBank under the accession [JBEBFO000000000](https://ncbi.nlm.nih.gov/nucl/JBEBFO000000000). The version described in this paper is [JBEBFO010000000](https://ncbi.nlm.nih.gov/nucl/JBEBFO010000000). Raw sequencing data has been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession [PRJNA1118351](https://ncbi.nlm.nih.gov/bioproject/PRJNA1118351). Genome assembly and annotations are also accessible at the JGI algal portal Phycosm (12): phycocosm.jgi.doe.gov/NanspC018_1/.

REFERENCES

1. Chiu SY, Kao CY, Tsai MT, Ong SC, Chen CH, Lin CS. 2009. Lipid accumulation and CO₂ utilization of *Nannochloropsis oculata* in response to CO₂ aeration. *Bioresour Technol* 100:833–838. <https://doi.org/10.1016/j.biortech.2008.06.061>

2. Mimouni V, Couzinet-Mossion A, Ulmann L, Wielgosz-Collin G. 2018. Lipids from microalgae, p 109–131. In *Microalgae in health and disease prevention*. Academic Press.
3. Guillard RR. 1975. Culture of phytoplankton for feeding marine invertebrates, p 29–60. In *Culture of marine invertebrate animals: proceedings—1st conference on culture of marine invertebrate animals greenport*. Springer US, Boston, MA.
4. Larivière D, Abueg L, Brajuka N, Gallardo-Alba C, Grüning B, Ko BJ, Ostrovsky A, Palmada-Flores M, Pickett BD, Rabbani K, et al. 2024. Scalable, accessible and reproducible reference genome assembly and evaluation in Galaxy. *Nat Biotechnol* 42:367–370. <https://doi.org/10.1038/s41587-023-02100-3>
5. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* 44:W3–W10. <https://doi.org/10.1093/nar/gkw343>
6. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
7. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 17:10. <https://doi.org/10.14806/ej.17.1.200>
8. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat Methods* 18:170–175. <https://doi.org/10.1038/s41592-020-01056-5>
9. 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>
10. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36:2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025>
11. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
12. Grigoriev IV, Hayes RD, Calhoun S, Kamel B, Wang A, Ahrendt S, Dusheyko S, Nikitin R, Mondo SJ, Salamov A, Shabalov I, Kuo A. 2021. PhycoCosm, a comparative algal genomics resource. *Nucleic Acids Res* 49:D1004–D1011. <https://doi.org/10.1093/nar/gkaa898>
13. Ohan JA, Hovde BT, Zhang XL, Davenport KW, Chertkov O, Han C, Twary SN, Starkenburg SR. 2019. Nuclear genome assembly of the microalga *Nannochloropsis salina* CCMP1776. *Microbiol Resour Announc* 8:10–1128. <https://doi.org/10.1128/MRA.00750-19>
14. Schwartz AS, Brown R, Ajjawi I, McCarren J, Atilla S, Bauman N, Richardson TH. 2018. Complete genome sequence of the model oleaginous alga *Nannochloropsis gaditana* CCMP1894. *Genome Announc* 6:10–1128. <https://doi.org/10.1128/genomeA.01448-17>
15. Corteggiani Carpinelli E, Telatin A, Vitulo N, Forcato C, D'Angelo M, Schiavon R, Vezzi A, Giacometti GM, Morosinotto T, Valle G. 2014. Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. *Mol Plant* 7:323–335. <https://doi.org/10.1093/mp/sst120>