


Nuclear Genome Assembly of the Microalga *Nannochloropsis salina* CCMP1776

J. A. Ohan,^a B. T. Hovde,^a X. L. Zhang,^a K. W. Davenport,^a O. Chertkov,^a C. Han,^a S. N. Twary,^a  S. R. Starkenburg^a

^aBioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

ABSTRACT *Nannochloropsis salina* is a halotolerant, high-lipid-producing microalga that is being explored as a biofuel production species. Here, we report an improved high-quality draft assembly and annotation for the nuclear genome of *N. salina* strain CCMP1776.

Nannochloropsis is a genus of eukaryotic microalgae (1) known for high lipid content and the ability to be maintained in large-volume outdoor cultures (2, 3). It can also produce auxiliary products such as the pigments astaxanthin, zeaxanthin, and canthaxanthin and the dietary supplement eicosapentaenoic acid (EPA) (4–7), an omega-3 fatty acid. Further, *Nannochloropsis* is tractable for genetic modification (8, 9), with evidence for homologous recombination in some strains (8). *Nannochloropsis salina* is a halotolerant strain known to accumulate 50 to 70% of its dry weight as lipid under nitrogen starvation (5, 10), making it an attractive candidate as a biofuel feedstock.

Nannochloropsis salina strain CCMP1776 was initially isolated in 1965 from Skate Point, Scotland (55.75°N, 4.96°W), and was deposited in the Bigelow culture collection in 1997.

CCMP1776 was cultivated in f/2 medium at room temperature under ~50 microeinsteins per meter squared per second and a 24-h light regime. Cultures growing under linear growth were harvested using centrifugation. Cells were lysed in AP1 buffer with a single pass through an Avestin Emulsiflex B-15 press at 30,000 lb/in². Genomic DNA was purified using the Qiagen DNeasy plant maxikit following the manufacturer's protocols.

Genomic DNA from CCMP1776 was sequenced and assembled using a combination of Illumina (11) and 454 (12) technologies. For this genome, we constructed and sequenced an Illumina GAII shotgun library, which generated 466 million reads totaling 85 Gb (90× coverage), and 2 paired-end 454 Titanium libraries with an average insert size of 5 kb, which generated 2,786,633 reads totaling 7.1 Gb of 454 data (16× coverage). The 454 Titanium standard data and the 454 paired-end data were assembled together with Newbler version 2.3 (091027_1459). The Newbler consensus sequences were computationally shredded into 10-kb overlapping fake reads (shreds) using an in-house script, resulting in 1.5× coverage of this assembly. Illumina sequencing data were assembled with Velvet version 1.0.13 (13), and the consensus sequence was computationally shredded into 10-kb overlapping shreds. We integrated the 454 Newbler consensus shreds, the Illumina Velvet consensus shreds, and the read pairs in the 454 paired-end library using parallel Phrap version 1.080812 (High Performance Software, LLC). Possible misassemblies were corrected using Gap Resolution (14) and Dupfinisher (15). The Gap Resolution software is available from the Department of Energy and the Lawrence Berkeley National Laboratory.

The final genome assembly was 27.6 Mbp contained in 194 scaffolds. The N_{50} value of this assembly is 828,788 bp, and the GC content is 54.88%. Genome annotation was performed using the BRAKER version 2 training and annotation pipeline (16) utilizing

Citation 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:e00750-19. <https://doi.org/10.1128/MRA.00750-19>.

Editor Antonis Rokas, Vanderbilt University

Copyright © 2019 Ohan 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 S. R. Starkenburg, shawns@lanl.gov.

Received 25 July 2019

Accepted 2 October 2019

Published 31 October 2019

254 million transcriptomic reads (paired end, 2 × 150 bp). Functional annotation of the 10,522 genes was performed using InterProScan 5 (17) and BLASTp searches against the UniProt (18) protein BLAST database. This genome will spur the continued development of algae for use as biofuel feedstock and provide prerequisite information needed for genetic manipulation.

Data availability. All sequences have been deposited in NCBI under BioSample number [SAMN10354914](https://www.ncbi.nlm.nih.gov/biosample/SAMN10354914) and GenBank accession number [SDOX00000000](https://www.ncbi.nlm.nih.gov/genbank/SDOX00000000). Genome assembly and annotations are also available at greenhouse.lanl.gov. The 454 raw sequencing data are available under NCBI SRA numbers [SRR9992831](https://www.ncbi.nlm.nih.gov/sra/SRR9992831) and [SRR9992828](https://www.ncbi.nlm.nih.gov/sra/SRR9992828). The Illumina raw reads are available under NCBI SRA number [SRR9992830](https://www.ncbi.nlm.nih.gov/sra/SRR9992830). The transcriptomic reads have been deposited under NCBI SRA number [SRR9992829](https://www.ncbi.nlm.nih.gov/sra/SRR9992829).

ACKNOWLEDGMENTS

This work was funded by the U.S. Department of Energy under contract NL0029949 to S.R.S. and contract DE-EE0003046 to the National Alliance for Advanced Biofuels and Bioproducts.

REFERENCES

- Hibberd D. 1981. Notes on the taxonomy and nomenclature of the algal classes Eustigmatophyceae and Tribophyceae (synonym Xanthophyceae). *Bot J Linn Soc* 82:93–119. <https://doi.org/10.1111/j.1095-8339.1981.tb00954.x>.
- Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC. 2012. Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana*. *Nat Commun* 3:686. <https://doi.org/10.1038/ncomms1688>.
- Rodolfi L, Chini Zittelli G, Bassi N, Padovani G, Biondi N, Bonini G, Tredici MR. 2009. Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol Bioeng* 102:100–112. <https://doi.org/10.1002/bit.22033>.
- Rocha JM, Garcia JE, Henriques MH. 2003. Growth aspects of the marine microalga *Nannochloropsis gaditana*. *Biomol Eng* 20:237–242. [https://doi.org/10.1016/S1389-0344\(03\)00061-3](https://doi.org/10.1016/S1389-0344(03)00061-3).
- Boussiba S, Vonshak A, Cohen Z, Avissar Y, Richmond A. 1987. Lipid and biomass production by the halotolerant microalga *Nannochloropsis salina*. *Biomass* 12:37–47. [https://doi.org/10.1016/0144-4565\(87\)90006-0](https://doi.org/10.1016/0144-4565(87)90006-0).
- Zou N, Zhang C, Cohen Z, Richmond A. 2000. Production of cell mass and eicosapentaenoic acid (EPA) in ultrahigh cell density cultures of *Nannochloropsis* sp. (Eustigmatophyceae). *Eur J Phycol* 35:127–133. <https://doi.org/10.1017/S0967026200002699>.
- Pal D, Khozin-Goldberg I, Cohen Z, Boussiba S. 2011. The effect of light, salinity, and nitrogen availability on lipid production by *Nannochloropsis* sp. *Appl Microbiol Biotechnol* 90:1429–1441. <https://doi.org/10.1007/s00253-011-3170-1>.
- Kilian O, Benemann CS, Niyogi KK, Vick B. 2011. High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp. *Proc Natl Acad Sci U S A* 108:21265–21269. <https://doi.org/10.1073/pnas.1105861108>.
- Jinkerson RE, Radakovits R, Posewitz MC. 2013. Genomic insights from the oleaginous model alga *Nannochloropsis gaditana*. *Bioengineered* 4:37–43. <https://doi.org/10.4161/bioe.21880>.
- Loira N, Mendoza S, Cortés MP, Rojas N, Travisany D, Di Genova A, Gajardo N, Ehrenfeld N, Maass A. 2017. Reconstruction of the microalga *Nannochloropsis salina* genome-scale metabolic model with applications to lipid production. *BMC Syst Biol* 11:66. <https://doi.org/10.1186/s12918-017-0441-1>.
- Bennett S. 2004. Solexa Ltd. *Pharmacogenomics J* 5:433–438. <https://doi.org/10.1517/14622416.5.4.433>.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. 2005. Genome sequencing in micro-fabricated high-density picolitre reactors. *Nature* 437:376. <https://doi.org/10.1038/nature03959>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <https://doi.org/10.1101/gr.074492.107>.
- Trong S, LaButti K, Foster B, Han C, Brettin T, Lapidus A. 2009. Gap Resolution: a software package for improving Newbler genome assemblies. LBNL report LBNL-1899E poster. <https://escholarship.org/uc/item/4vc652xh>.
- Han C, Chain P. 2006. Finishing repetitive regions automatically with Dupfinisher, p 142–147. *In* Arabnia HR, Valafar H (ed), *Proceedings of the 2006 International Conference on Bioinformatics & Computational Biology, BIOCOMP'06*. CSREA Press, Las Vegas, NV.
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32:767–769. <https://doi.org/10.1093/bioinformatics/btv661>.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30: 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
- The UniProt Consortium. 2018. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 46:2699. <https://doi.org/10.1093/nar/ky092>.