




Draft Nuclear Genome Sequence of the Halophilic and Beta-Carotene-Accumulating Green Alga *Dunaliella salina* Strain CCAP19/18

Juergen E. W. Polle,^{a,b} Kerrie Barry,^c John Cushman,^d Jeremy Schmutz,^{c,e} Duc Tran,^a Leyla T. Hathwaik,^d Won C. Yim,^d Jerry Jenkins,^e  Zaid McKie-Krisberg,^a Simon Prochnik,^{c,*} Erika Lindquist,^c Rhyhan B. Dockter,^c Catherine Adam,^c Henrik Molina,^f Jakob Bunkenborg,^g EonSeon Jin,^h Mark Buchheim,ⁱ Jon Magnuson^j

Department of Biology, Brooklyn College of the City University of New York, Brooklyn, New York, USA^a; The Graduate Center of the City University of New York, New York, New York, USA^b; Department of Energy, Joint Genome Institute, Walnut Creek, California, USA^c; University of Nevada, Department of Biochemistry and Molecular Biology, Reno, Nevada, USA^d; HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA^e; The Proteomics Resource Center, The Rockefeller University, New York, New York, USA^f; Alphalyse A/S, Odense, Denmark^g; Department of Life Science, Hanyang University, Research Institute for Natural Sciences, Seoul, Republic of Korea^h; Biological Science, The University of Tulsa, Tulsa, Oklahoma, USAⁱ; Pacific Northwest National Laboratory, Richland, Washington, USA^j

ABSTRACT The halotolerant alga *Dunaliella salina* is a model for stress tolerance and is used commercially for production of beta-carotene (=pro-vitamin A). The presented draft genome of the genuine strain CCAP19/18 will allow investigations into metabolic processes involved in regulation of stress responses, including carotenogenesis and adaptations to life in high-salinity environments.

The halotolerant green alga *Dunaliella salina* Teodoresco is the type species of the *Dunaliella* genus (class *Chlorophyceae*) (1, 2). The biflagellate cells grow best at salinities of about 1.5 M (3). Often, new isolates of halotolerant algae of this general appearance are incorrectly identified as *D. salina*. Strain CCAP 19/18 is a genuine representative of the species *D. salina* (4), making it a reference for all ongoing and future studies on this and other *Dunaliella* species.

Due to the absence of a rigid cell wall and the presence of a glycocalyx (5), *D. salina* cells can adjust rapidly to large changes in salinity. Thus, *D. salina* is a model for studies of salinity tolerance (6–11). Further, under abiotic stress cells turn orange due to overaccumulation of beta-carotene in plastidic oil globules (12). Beta-carotene may accumulate to over 8% of the cellular dry weight, making the alga an important crop for the health food market for production of natural beta-carotene (13–17). With the recent surge in *Dunaliella* “omics” research and currently 19 BioProjects deposited in NCBI (NIH), we offer this early release draft genome sequence of strain CCAP19/18 to support and encourage ongoing and future research.

Following genomic DNA extraction from cells, libraries were generated for sequencing using either the Illumina or the PacBio platforms, overall producing 554.53 Gbp (after quality filtering). The tight-insert protocol was used for library construction for the Illumina HiSeq 2500 paired-end (PE) runs, which produced 219.9 Gbp (after quality filtering). For Illumina HiSeq 2000 sequencing preparation, the ligation-free paired end (LFPE) construction protocol was used for four libraries, the Illumina standard PE unamplified protocol was used for three libraries, and the Illumina CLIP PE protocol was used for one library, in total providing 341.6 Gbp (after quality filtering). Sequences generated on the Illumina platform provided 62× coverage and those

Received 21 September 2017 Accepted 25 September 2017 Published 26 October 2017

Citation Polle JEW, Barry K, Cushman J, Schmutz J, Tran D, Hathwaik LT, Yim WC, Jenkins J, McKie-Krisberg Z, Prochnik S, Lindquist E, Dockter RB, Adam C, Molina H, Bunkenborg J, Jin E, Buchheim M, Magnuson J. 2017. Draft nuclear genome sequence of the halophilic and beta-carotene-accumulating green alga *Dunaliella salina* strain CCAP19/18. *Genome Announc* 5:e01105-17. <https://doi.org/10.1128/genomeA.01105-17>.

Copyright © 2017 Polle 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 Juergen E. W. Polle, jpolle@brooklyn.cuny.edu.

* Present address: Simon Prochnik, Intrexon Corporation, South San Francisco, California, USA.

from paired-end reads 27× (from 2-kb, 4-kb, and 6-kb libraries). Additionally, 16 PacBio single-molecule real-time (SMRT) libraries were constructed and sequenced on the PacBio RS platform, providing, after quality filtering, a total of 3.162 Gbp and resulting in about 10× coverage. The chloroplast and mitochondrial sequences (18) were removed prior to assembly.

The genome was assembled with ALLPATHS-LG (19). The resulting assembly was scaffolded using SSPACE (20) with three fosmid libraries (total of 0.50× read coverage). The final assembly was screened for contaminants against the NCBI nonredundant database. The *D. salina* v1.0 draft genome is provided through the Phytozome portal from the U.S. Department of Energy (DOE) Joint Genome Institute (JGI) at <http://phytozome.jgi.doe.gov/>. The main genome assembly provides approximately 343.7 Mb arranged in 5,512 scaffolds ($N_{50} = 353.0$ kbp) and a 49.1 G+C content. The contig N_{50} ($L_{50} = 10,635$ (7.8 kb)), and 1,241 scaffolds are >50 kb in size, representing approximately 92.0% of the genome. With a conservative annotation approach using a repeat-masked assembly, the presented draft genome contains 16,697 loci providing 18,801 protein-coding transcripts. A less conservative annotation approach without masking repeats revealed 36,851 protein-coding genes. The assembly and annotation are expected to provide the foundation for future studies on halotolerance, secondary carotenoid accumulation, and algal stress biology.

Accession number(s). The whole-genome sequencing project was deposited in GenBank under BioProject number PRJNA32771. The project BioSample number is SAMN02746051 (sample name 1014865, SRA number SRS1543949). A trace file archive is available under NCBI project ID 32771. The draft genome sequence is accessible under GenBank accession number [NSFN00000000](https://www.ncbi.nlm.nih.gov/nuccore/NSFN00000000). This paper describes the first version, NSF01000000.

ACKNOWLEDGMENTS

The work conducted by the U.S. Department of Energy (DOE) Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC02-05CH11231. This research was supported, in part, under the National Science Foundation Grants CNS-0958379, CNS-0855217, and ACI-1126113 and the City University of New York High Performance Computing Center at the College of Staten Island.

REFERENCES

- Teodoresco EC. 1905. Organization et developpement du *Dunaliella*, nouveau genre de Volvocacee-Polyblepharidee. *Beih Bot Zentralblatt Band 18 Abteilung 1*:215–232.
- Hamburger C. 1905. Zur Kenntnis der *Dunaliella salina* und einer Amöbe aus Salinenwasser von Cagliari. *Archiv Protistenkund* 6:111–130.
- Preisig HR. 1992. Chapter 1. Morphology and taxonomy, p 1–15. *In* Ben-Amotz A, Avron M (ed), *Dunaliella: physiology, biochemistry, and biotechnology*. CRC Press, Boca Raton, FL.
- Gonzalez MA, Gomez PI, Polle JEW. 2009. Chapter 2. Taxonomy and phylogeny of the genus *Dunaliella*, p 15–44. *In* Ben-Amotz A, Polle JEW, Subba Rao DV (ed), *The alga Dunaliella: biodiversity, physiology, genomics and biotechnology*. CRC Press, Enfield, NH.
- Borowitzka MA, Borowitzka LJ. 1988. Algal growth media and sources of cultures, p 456–465. *In* Borowitzka MA, Borowitzka LJ (ed), *Micro-algal biotechnology*. Cambridge University Press, Cambridge, UK.
- Avron M. 1992. Chapter 6. Osmoregulation, p 135–159. *In* Ben-Amotz A, Avron M (ed), *Dunaliella: physiology, biochemistry, and biotechnology*. CRC Press, Boca Raton, FL.
- Borowitzka LJ, Kessly DS, Brown AD. 1977. The salt relations of *Dunaliella*. Further observations on glycerol production and its regulation. *Arch Microbiol* 113:131–138. <https://doi.org/10.1007/BF00428592>.
- Chen H, Jiang JG. 2009. Osmotic responses of *Dunaliella* to the changes of salinity. *J Cell Physiol* 219:251–258. <https://doi.org/10.1002/jcp.21715>.
- Ginzburg M. 1988. *Dunaliella*: a green alga adapted to salt. *Adv Bot Res* 14:93–183. [https://doi.org/10.1016/S0065-2296\(08\)60271-2](https://doi.org/10.1016/S0065-2296(08)60271-2).
- Oren A. 2017. Glycerol metabolism in hypersaline environments. *Environ Microbiol* 19:851–863. <https://doi.org/10.1111/1462-2920.13493>.
- Wegmann K. 1971. Osmotic regulation of photosynthetic glycerol production in *Dunaliella*. *Biochim Biophys Acta* 234:317–323. [https://doi.org/10.1016/0005-2728\(71\)90197-6](https://doi.org/10.1016/0005-2728(71)90197-6).
- Ben-Amotz A, Katz A, Avron M. 1982. Accumulation of β -carotene in halotolerant algae: purification and characterization of β -carotene-rich globules from *Dunaliella bardawil* (*Chlorophyceae*). *J Phycol* 18:529–537.
- Borowitzka LJ, Borowitzka MA, Moulton TP. 1984. The mass culture of *Dunaliella salina* for fine chemicals: from laboratory to pilot plant. *Hydrobiologia* 116–117:115–121. <https://doi.org/10.1007/BF00027649>.
- Borowitzka LJ. 1991. Development of western biotechnology's algal β -carotene plant. *Bioresour Technol* 38:251–252. [https://doi.org/10.1016/0960-8524\(91\)90164-F](https://doi.org/10.1016/0960-8524(91)90164-F).
- Ben-Amotz A, Shaish A, Avron M. 1991. The biotechnology of cultivating *Dunaliella* for production of β -carotene rich algae. *Bioresour Technol* 38:233–235. [https://doi.org/10.1016/0960-8524\(91\)90160-L](https://doi.org/10.1016/0960-8524(91)90160-L).
- Ben-Amotz A. 2004. Chapter 13. Industrial production of microalgal cell-mass and secondary products—major industrial species: *Dunaliella*, p 273–280. *In* Richmond A (ed), *The handbook of microalgal culture: biotechnology and applied phycology*. Wiley-Blackwell, Hoboken, NJ. <https://doi.org/10.1002/9780470995280.ch13>.
- Borowitzka MA. 2013. *Dunaliella*: biology, production, and markets, p 359–368. *In* Richmond A, Hu Q (ed), *The handbook of microalgal culture: applied phycology and biotechnology*, 2nd ed. Wiley-Blackwell, Hoboken, NJ.
- Smith DR, Lee RW, Cushman JC, Magnuson JK, Tran D, Polle JEW. 2010.

- The *Dunaliella salina* organelle genomes: large sequences, inflated with intronic and intergenic DNA. *BMC Plant Biol* 10:83. <https://doi.org/10.1186/1471-2229-10-83>.
19. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* 108:1513–1518. <https://doi.org/10.1073/pnas.1017351108>.
20. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.