



Draft Nuclear Genome Sequence of the Liquid Hydrocarbon–Accumulating Green Microalga *Botryococcus braunii* Race B (Showa)

Daniel R. Browne,^a Jerry Jenkins,^b Jeremy Schmutz,^{b,c} Shengqiang Shu,^c Kerrie Barry,^c Jane Grimwood,^b Jennifer Chiniquy,^{c*} Aditi Sharma,^c Thomas D. Niehaus,^{d*} Taylor L. Weiss,^{a*} Andrew T. Koppisch,^e David T. Fox,^f Suraj Dhungana,^{f*} Shigeru Okada,^{g,h} Joe Chappell,^d Timothy P. Devarenne^a

Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas, USA^a; HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA^b; Joint Genome Institute, Walnut Creek, California, USA^c; Plant Biology Program and Department of Pharmaceutical Sciences, University of Kentucky, Lexington, Kentucky, USA^d; Department of Chemistry, Northern Arizona University, Flagstaff, Arizona, USA^e; Chemistry Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA^f; Laboratory of Aquatic Natural Products Chemistry, Graduate School of Agricultural and Life Sciences, the University of Tokyo, Bunkyo, Tokyo, Japan^g; Japan Science and Technology Agency, Core Research for Evolutional Science and Technology (CREST), Chiyoda, Tokyo, Japan^h

ABSTRACT *Botryococcus braunii* has long been known as a prodigious producer of liquid hydrocarbon oils that can be converted into combustion engine fuels. This draft genome for the B race of *B. braunii* will allow researchers to unravel important hydrocarbon biosynthetic pathways and identify possible regulatory networks controlling this unusual metabolism.

The oil-producing traits of the colony-forming green microalga *Botryococcus braunii* have attracted study since the 19th century (1). In the late 20th century, morphologically similar strains were differentiated into three chemically distinct “races” (A, B, L) defined by the types of oils biosynthesized (alkadienes, botryococcenes, lycopadiene) in each race (2). *B. braunii* remnants have been identified in organic sediments dating from the Precambrian to the Permian period (3), indicating that *B. braunii* contributed material to petroleum formations (4). Like petroleum, *B. braunii* oil can be catalytically cracked, yielding fuel-range distillates (5, 6). Renewable fuel interests are currently driving *B. braunii* genomic research, and here we present an early release draft genome of the *B. braunii* race B (Showa) strain, which has an estimated size of 166.2 ± 2.2 Mb (7).

Genomic DNA was extracted and used to construct four Illumina libraries. First, a 2 × 250-bp paired-end library constructed from 800-bp fragments was sequenced to 700× coverage on an Illumina HiSeq 2500 platform. Second and third, 2 × 150-bp mate-pair libraries with 1.5-kb and 4-kb inserts were sequenced to 200× and 150× coverage on a HiSeq 2000. Fourth, a 2 × 300-bp mate-pair library with a 15-kb insert was sequenced to 3× coverage on an Illumina MiSeq. Two PacBio SMRTbell libraries were also constructed and sequenced to 200× coverage on the PacBio RS II platform.

The PacBio data were assembled with FALCON-Unzip, and the resulting sequences were polished using Quiver (8). To detect misassemblies, the 15-kb library was aligned to the sequences and clone coverage at each base computed. Nineteen misassemblies were detected and broken. The Illumina data were assembled with DISCOVAR. Sequences were identified in the DISCOVAR assembly that were not present in the FALCON assembly. The DISCOVAR assembly was masked using 24-mers from the FALCON assembly, and 487 unmasked sequences (1.396 Mb) were extracted. These sequences were combined with the

Received 22 February 2017 Accepted 27 February 2017 Published 20 April 2017

Citation Browne DR, Jenkins J, Schmutz J, Shu S, Barry K, Grimwood J, Chiniquy J, Sharma A, Niehaus TD, Weiss TL, Koppisch AT, Fox DT, Dhungana S, Okada S, Chappell J, Devarenne TP. 2017. Draft nuclear genome sequence of the liquid hydrocarbon–accumulating green microalga *Botryococcus braunii* race B (Showa). Genome Announc 5:e00215-17. <https://doi.org/10.1128/genomeA.00215-17>.

Copyright © 2017 Browne 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 Timothy P. Devarenne, tpd8@tamu.edu.

* Present address: Jennifer Chiniquy, Lawrence Berkeley National Laboratory, Berkeley, California, USA; Thomas D. Niehaus, Horticultural Sciences Department, University of Florida, Gainesville, Florida, USA; Taylor L. Weiss, MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan, USA; Suraj Dhungana, Waters Corporation, Milford, Massachusetts, USA. D.R.B. and J.J. contributed equally to this work.

broken FALCON assembly and scaffolded with the 15-kb library using SSPACE (9). Finally, the assembly was error-corrected using the Illumina data. Analysis revealed 523 scaffolds (19.8 Mb) that did not share a significant number of 24-mers with the rest of the assembly. These sequences were aligned to the NCBI NR database, identified as prokaryotic contamination, and removed from the assembly. Mitochondrial and chloroplast sequences (10) were removed prior to the assembly.

The final draft assembly consists of 184,385,342 bp in 2,752 scaffolds ($N_{50} = 373$ kb) with 49.6% G+C content and 1,148 gaps (4.611 Mb). There are 18,726 predicted genes with a mean of 5.7 exons per gene, a median exon length of 178 bp, and a median intron length of 578 bp. The 1,437 scaffolds with no genic content (the largest is 49,840 bp) account for 6,183,350 bp. This assembly provides a strong basis for functional and comparative analyses and will help elucidate the genetic basis of oil metabolism in *B. braunii*.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number [MVGU00000000](https://doi.org/10.1093/genomea/00524-16). The version described in this paper is the first version, MVGU01000000.

ACKNOWLEDGMENTS

We thank David Kudrna, University of Arizona, for high-molecular-weight DNA extraction used for PacBio sequencing. This work was supported by the JGI 2010 Community Sequencing Program, grant no. CSP2010-784140 to A.T.K., T.P.D., J.C., O.K., and D.T.F., and by NSF grant CHE-1412648 to A.T.K. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

REFERENCES

- Blackburn KB. 1936. *Botryococcus* and the algal coals part I: a reinvestigation of the alga *Botryococcus braunii*. *Trans Roy Soc Edin* 58:841–854.
- Metzger P, Largeau C. 2005. *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids. *Appl Microbiol Biotechnol* 66: 486–496. <https://doi.org/10.1007/s00253-004-1779-z>.
- Glikson M, Lindsay K, Saxby J. 1989. *Botryococcus*—a planktonic green alga, the source of petroleum through the ages: transmission electron microscopical studies of oil shales and petroleum source rocks. *Org Geochem* 14:595–608. [https://doi.org/10.1016/0146-6380\(89\)90039-9](https://doi.org/10.1016/0146-6380(89)90039-9).
- Rodriguez ND, Paul Philp R. 2012. Productivity and paleoclimatic controls on source rock character in the Aman Trough, north central Sumatra, Indonesia. *Org Geochem* 45:18–28. <https://doi.org/10.1016/j.orggeochem.2012.01.004>.
- Tracy NI, Crunkleton DW, Price GL. 2011. Catalytic cracking of squalene to gasoline-range molecules. *Biomass Bioenerg* 35:1060–1065. <https://doi.org/10.1016/j.biombioe.2010.11.027>.
- Hillen LW, Pollard G, Wake LV, White N. 1982. Hydrocracking of the oils of *Botryococcus braunii* to transport fuels. *Biotechnol Bioeng* 24:193–205. <https://doi.org/10.1002/bit.260240116>.
- Weiss TL, Spencer Johnston J, Fujisawa K, Sumimoto K, Okada S, Chappell J, Devarenne TP. 2010. Phylogenetic placement, genome size, and GC content of the liquid-hydrocarbon-producing green microalga *Botryococcus braunii* strain Berkeley (Showa) (Chlorophyta). *J Phycol* 46: 534–540. <https://doi.org/10.1111/j.1529-8817.2010.00820.x>.
- Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC. 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* 13:1050–1054. <https://doi.org/10.1038/nmeth.4035>.
- 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>.
- Blifernez-Klassen O, Wibberg D, Winkler A, Blom J, Goesmann A, Kalinowski J, Kruse O. 2016. Complete chloroplast and mitochondrial genome sequences of the hydrocarbon oil-producing green microalga *Botryococcus braunii* race B (Showa). *Genome Announc* 4(3):e00524-16. <https://doi.org/10.1128/genomeA.00524-16>.