AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

High-Quality Draft Genome Sequence of the Green Alga *Tetraselmis striata* (Chlorophyta) Generated from PacBio Sequencing

Christina R. Steadman Tyler,[a] Blake T. Hovde,[a] Hajnalka E. Daligault,[a] Xiang Li Zhang,[a] Yuliya Kunde,[a] Babetta L. Marrone,[a] Scott N. Twary,[a] Shawn R. Starkenburg[a]

[a]Los Alamos National Laboratory, Biosciences Division, Los Alamos, New Mexico, USA

ABSTRACT A high-quality draft genome sequence of the microalgal species *Tetraselmis striata* was generated using PacBio sequencing. The assembled genome is 228 Mb, derived from 3,613 polished contigs at 84× coverage depth. This genome contains an average GC content of 57.9% and 48,906 predicted genes.

Photosynthetic eukaryotic microalgae sequester carbon into biomolecules that can be used for alternative sources of biomass and bioenergy. There is concerted effort among researchers from government, academia, and industry to identify and characterize algal species with desired feedstock phenotypes. *Tetraselmis* species comprise a genus of large, unicellular green algae that have great biofuel potential due to their adaptability to variable environmental conditions (1), symbiotic relationships with an extensive phycosphere (2), and ability to sequester carbon into lipids under stress (3). These marine algae are approximately 6 to 10 μm in diameter and covered in theca (sugar acid scales), with four flagella for motility (4, 5). Here, we report the first genome from the species *Tetraselmis striata* (formerly *Platymonas*).

Nannochloropsis salina CCMP1776 was grown in outdoor production ponds in southwestern Colorado (6). Samples collected from these ponds were cultured *in vitro*, and mixed algal species were noted. *Tetraselmis striata* was isolated from the *N. salina* culture using flow cytometry via size exclusion (10 μm versus 2 μm, respectively). *T. striata* was cultivated in modified f/2 medium (8.8 mM sodium nitrate) in 1-liter spin flasks and maintained at 25°C with 800 μmol · m^{-2} · s^{-1} fluorescent lighting on a 16h/8h light/dark cycle (7). Genomic DNA was extracted using agarose plugs and subsequently fragmented and purified to an average size of 20 to 40 kb for long-read sequencing. One PacBio SMRTbell library was generated. Eighteen single-molecule real-time (SMRT) cells were sequenced on the PacBio RS II platform using MagBead loading and C4 chemistry, and two SMRT cells were sequenced on the PacBio Sequel system using diffusion loading and C2 chemistry. *De novo* assembly was completed using Fast Alignment and CONsensus (FALCON) on sequencing data derived from the 20 SMRT cells (NCBI SRA accession no. SRR10043439 to SRR10043457). Based on previous transcriptomic data, the input genome size was set to 300 Mb, and seed coverage was set to 55×. There were 4,009,867 filtered subreads with a filtered subread length (N_{50}) of 10,277. Default parameters were used to clean and filter the data. The FALCON assembly was polished twice using ReSequencing via SMRT Link v5.0.0 and corrected by PILON v1.22 using Illumina reads (NCBI SRA accession no. SRR10043458); both tools were run under default parameters. The final assembled genome contains 227,954,216 bases derived from 3,613 contigs, ranging from 710 to 4,140,848, with a GC content of 57.9% and an N_{50} value of 126,807 bp. The resulting assembly was annotated using BRAKER v2.0 with the default parameters specified for transcriptome sequencing (RNA-seq)-based training. The *Tetraselmis* sp. RNA-seq data used for train-

Citation Steadman Tyler CR, Hovde BT, Daligault HE, Zhang XL, Kunde Y, Marrone BL, Twary SN, Starkenburg SR. 2019. High-quality draft genome sequence of the green alga *Tetraselmis striata* (Chlorophyta) generated from PacBio sequencing. Microbiol Resour Announc 8:e00780-19. https://doi.org/10.1128/MRA.00780-19.

Editor Jason E. Stajich, University of California, Riverside

Address correspondence to Christina R. Steadman Tyler, crt@lanl.gov.

Received 28 June 2019
Accepted 2 October 2019
Published 24 October 2019

ing is located in the NCBI SRA database (accession no. SRR1296733 from the The Marine Microbial Eukaryote Transcriptome Sequencing Project) (8). Annotation generated 48,906 predicted gene models.

Data availability. The annotated genome sequence is publicly available in the Greenhouse Knowledgebase at Los Alamos National Laboratory (https://greenhouse.lanl.gov/greenhouse) under "Algal Genomes: *Tetraselmis striata*" (https://greenhouse.lanl.gov/greenhouse/organisms/). This project has been deposited in GenBank under BioProject no. PRJNA542153, with the genome assembly deposited under accession no. VCJN00000000. The version described in this paper is the first version, VCJN01000000.

ACKNOWLEDGMENTS

This work was funded by the United States Department of Energy (DOE) Bioenergy Technologies Office (BETO) under Annual Operating Plan projects NL0026328 and NL0029949 and by a Los Alamos National Laboratory Director's Postdoctoral Fellowship (20170690PRD4 to C. R. Steadman Tyler).

REFERENCES

1. Pagarete A, Grébert T, Stepanova O, Sandaa RA, Bratbak G. 2015. Tsv-N1: a novel DNA algal virus that infects *Tetraselmis striata*. Viruses 7:3937–3953. https://doi.org/10.3390/v7072806.
2. Patidar SK, Kim SH, Kim JH, Park J, Park BS, Han MS. 2018. *Pelagibaca bermudensis* promotes biofuel competence of *Tetraselmis striata* in a broad range of abiotic stressors: dynamics of quorum-sensing precursors and strategic improvement in lipid productivity. Biotechnol Biofuels 11:1–16. https://doi.org/10.1186/s13068-018-1097-9.
3. Adarme-Vega TC, Thomas-Hall SR, Lim DKY, Schenk PM. 2014. Effects of long chain fatty acid synthesis and associated gene expression in microalga *Tetraselmis* sp. Mar Drugs 12:3381–3398. https://doi.org/10.3390/md12063381.
4. Gödel S, Becker B, Melkonian M. 2000. Flagellar membrane proteins of *Tetraselmis striata* butcher (Chlorophyta). Protist 151:147–159. https://doi.org/10.1078/1434-4610-00015.
5. Becker B, Melkonian M, Kamerling JP. 1998. The cell wall (theca) of *Tetraselmis striata* (Chlorophyta): macromolecular composition and structural elements of the complex polysaccharides. J Phycol 34:779–787. https://doi.org/10.1046/j.1529-8817.1998.340779.x.
6. 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.
7. Steadman Tyler CR, Sanders CK, Erickson RS, Dale T, Twary SN, Marrone BL. 2019. Functional and phenotypic flow cytometry characterization of *Picochlorum soloecismus*. Algal Res 43:101614. https://doi.org/10.1016/j.algal.2019.101614.
8. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A, Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, Nadathur G, Nagai S, Ngam PB, Palenik B, Pawlowski J, Petroni G, Piganeau G, Posewitz MC, Rengefors K, Romano G, Rumpho ME, Rynearson T, Schilling KB, Schroeder DC, Simpson AGB, Slamovits CH, Smith DR, Smith GJ, Smith SR, Sosik HM, Stief P, Theriot E, Twary SN, Umale PE, Vaulot D, Wawrik B, Wheeler GL, Wilson WH, Xu Y, Zingone A, Worden AZ. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. PLoS Biol 12:e1001889. https://doi.org/10.1371/journal.pbio.1001889.