Title
Annotated Genome Sequence of the High-Biomass-Producing Yellow-Green Alga Tribonema minus

Journal
Microbiology Resource Announcements, 10(24)

ISSN
2169-8287

Authors
Mahan, Kristina M
Polle, Jürgen EW
McKie-Krisberg, Zaid
et al.

Publication Date
2021-06-17

DOI
10.1128/mra.00327-21

Peer reviewed

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Annotated Genome Sequence of the High-Biomass-Producing Yellow-Green Alga *Tribonema minus*

Kristina M. Mahan,[a] Jürgen E. W. Polle,[b,c,d,f] Zaid McKie-Krisberg,[e] Anna Lipzen,[f] Alan Kuo,[f] Igor V. Grigoriev,[f,g] Todd W. Lane,[a] Aubrey K. Davis[d,h]

[a]Systems Biology, Sandia National Laboratories, Livermore, California, USA
[b]Department of Biology, Brooklyn College of the City University of New York, Brooklyn, New York, USA
[c]The Graduate Center of the City University of New York, New York, New York, USA
[d]MicroBio Engineering Inc., San Luis Obispo, California, USA
[e]Division of Information Services, SUNY Downstate Health Services University, Brooklyn, New York, USA
[f]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[g]Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, California, USA
[h]Civil and Environmental Engineering Department, California Polytechnic State University, San Luis Obispo, California, USA

ABSTRACT Here, we report the annotated genome sequence for a heterokont alga from the class Xanthophyceae. This high-biomass-producing strain, *Tribonema minus* UTEX B 3156, was isolated from a wastewater treatment plant in California. It is stable in outdoor raceway ponds and is a promising industrial feedstock for biofuels and bioproducts.

A draft haploid 158.35-Mb genome sequence for *Tribonema minus* strain UTEX B 3156 was assembled into 557 contigs containing 18,290 predicted protein-coding genes. *Tribonema* species are common to many freshwater and wastewater ecosystems and are distinguished by their filamentous, nonbranching, H-shaped bipartite walls (1). Some species can be high lipid and carbohydrate producers (2–12), making these organisms potential candidates for biodiesel production (2). In addition, these strains can be harvested without chemical flocculants and have applications in bioremediation of toxic compounds (13, 14). *T. minus* strain UTEX B 3156 was originally isolated from wastewater treatment ponds in San Luis Obispo, CA, and identified based on the cell morphology as well as on the ribosomal DNA (rDNA) sequence identity (15).

T. minus was grown photoautotrophically in bubble columns in 800 ml of BG11 medium (16) under fluorescent lighting at 100 μmol/m^2 s^1 at room temperature for 4 to 5 days. Genomic DNA was extracted by exposing agarose-embedded cells to cellulolytic enzymes as previously described (17). Then, 50 ml of culture was washed and resuspended in buffer (200 mM NaCl, 100 mM EDTA, 10 mM Tris [pH 7.2]), and 500 μl of the resuspended culture was mixed with premelted 1% low-melting-point agarose and distributed into plug molds (Bio-Rad, Hercules, CA). The plugs were allowed to solidify at 4°C and incubated in 50 ml of protoplasting solution (4% hemicellulase, 2% drielase, 0.1 mM sodium citrate, 1 M sorbitol, 240 mM EDTA, 10 mM β-mercaptoethanol) with shaking at 120 rpm, overnight at 37°C. The plugs were drained from the solution and incubated in 5 ml of lysis solution (2 mg/ml of proteinase K; 0.5 M EDTA, pH 9.5; 1% lauroyl sarcosine sodium salt) with shaking at 40 rpm, overnight at 50°C. The plugs were drained from the lysis solution and washed 3 times with Tris-EDTA (TE), pH 8.0 (10 mM Tris-HCl [pH 7.5] plus 1 mM EDTA [pH 8.0]), under gentle rocking. The plugs were warmed to 70°C for 7 min, added to 200 μl of prewarmed β-agarase solution (192 μl of TE [pH 8.0] plus 8 μl of β-agarase) (New England BioLabs [NEB], Ipswich, MA), and incubated for 16 h at 42°C. The genomic DNA was quality checked by running on a gel and using the Qubit 2.0 fluorometer (Invitrogen, Carlsbad, CA). Sequencing was

Citation Mahan KM, Polle JEW, McKie-Krisberg Z, Lipzen A, Kuo A, Grigoriev IV, Lane TW, Davis AK. 2021. Annotated genome sequence of the high-biomass-producing yellow-green alga *Tribonema minus*. Microbiol Resour Announc 10:e00327-21. https://doi.org/10.1128/MRA.00327-21.

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Aubrey K. Davis, AubreyDavis@MicrobioEngineering.com.

Received 12 April 2021
Accepted 25 May 2021
Published 17 June 2021

performed by Genewiz (South San Francisco, CA, USA). A 20-kb PacBio (Menlo Park, CA, USA) SMRTbell library was prepared using the BluePippin size selection system (Sage Science, Beverly, MA, USA) per the manufacturer's protocol. Two single-molecule real-time (SMRT) cells were sequenced and collectively produced 912,479 subreads with a mean subread length of 6,675 bp. This result provided 24,273 Mb of data, which was approximately 121× coverage of the assembled genome size (18). The PacBio reads were quality assessed via the error-correction step of the Canu v2.1.1 assembler, and subreads greater than 5 kb in length were assembled using Canu v2.1.1 (correctedErrorRate=0.085 corMinCoverage=0 corMhapSensitivity=high) (19). The Nextera XT DNA library preparation kit for Illumina was used for target enrichment DNA library preparation following the manufacturer's recommendations (San Diego, CA, USA). The additional Illumina HiSeq X Ten platform sequencing (2 × 150 bp) produced 141,827,758 reads, totaling 42,548 Mb, with a mean quality score of 35.98 and 94.13% bases having quality scores of ≥30. The Illumina paired-end sequencing reads were preprocessed using AfterQC v0.9.7 (20) and used to polish the Canu assembly with Pilon v1.23 (21). Using BWA-MEM v0.7.17 (22), 92.2% of the Illumina reads were mapped onto the assembled reference genome. The chloroplast and mitochondrial genome sequences were assembled using Fast-Plast v1.2.8 (23) and NOVOPlasty v4.2 (24). Default parameters were used except where otherwise noted.

T. minus RNA was extracted from pooled cells grown under various growth conditions in bubble columns (nitrogen depleted, low/high density, low/high light, early/late growth phase), using the RNeasy extraction kit from Qiagen. The RNA library preparations and sequencing reactions were conducted at Genewiz, LLC (South Plainfield, NJ, USA). The RNA samples were quantified using the Qubit 2.0 fluorometer (Invitrogen), and the RNA integrity was checked using the TapeStation 4200 platform (Agilent Technologies, Palo Alto, CA, USA). RNA sequencing libraries were prepared using the NEBNext Ultra RNA library prep kit for Illumina using the manufacturer's instructions (NEB). Briefly, mRNAs were initially enriched with oligo(dT) beads. The enriched mRNAs were fragmented for 15 min at 94°C. First-strand and second-strand cDNAs were subsequently synthesized. cDNA fragments were end repaired and adenylated at the 3′ ends, and universal adapters were ligated to the cDNA fragments, followed by index addition and library enrichment using PCR with limited cycles. The sequencing library was validated on the Agilent TapeStation platform and quantified using the Qubit 2.0 fluorometer (Invitrogen), as well as quantitative PCR (KAPA Biosystems, Wilmington, MA, USA). rRNA depletion was performed using the Ribo-Zero rRNA removal kit (Illumina). RNA sequencing libraries were prepared using the NEBNext Ultra RNA library prep kit for Illumina following the manufacturer's recommendations (NEB). Briefly, enriched RNAs were fragmented for 15 min at 94°C. First-strand and second-strand cDNAs were subsequently synthesized. cDNA fragments were end repaired and adenylated at the 3′ ends, and universal adapters were ligated to the cDNA fragments, followed by index addition and library enrichment with limited-cycle PCR. The sequencing libraries were validated using the Agilent TapeStation 4200 platform and quantified using the Qubit 2.0 fluorometer (Invitrogen) as well as quantitative PCR (Applied Biosystems, Carlsbad, CA, USA).

The sequencing libraries were clustered on a single lane of a flow cell. After clustering, the flow cell was loaded onto the Illumina HiSeq instrument (4000 or equivalent) according to the manufacturer's instructions. The samples were sequenced using a 2 × 150-bp paired-end (PE) configuration. Image analysis and base calling were conducted using the HiSeq control software (HCS). The raw sequence data (BCL files) generated using the Illumina HiSeq instrument were converted into fastq files and demultiplexed using Illumina's bcl2fastq v2.17 software. One mismatch was allowed for index sequence identification. Transcriptome sequencing (RNA-Seq) was carried out by Genewiz using the Illumina HiSeq platform (2 × 150 bp), which produced 132.88 Mb of reads with a mean quality score of 38.07 and 91.27% of bases having a quality score of ≥30. Sequencing yielded 39,864 Mb. The transcriptome was assembled using Trinity (25).

The assembled genome and transcriptome were used as inputs for the U.S. Department

TABLE 1 Genome assembly and annotation statistics of *T. minus* strain UTEX B 3156

Feature	Statistic
Estimated genome assembly size (Mb)	158.35
No. of contigs	557
N_{50} (bp)	768,631
L_{50}	66
Largest scaffold (Mb)	2.45
GC content (%)	56.96
Telomere repeat sequence	TTAGGG
No. of gene models	18,290
Avg gene length (bp)	5,210
Chloroplast length (bp)	136,609
Mitochondrion length (bp)	44,644

of Energy Joint Genome Institute (JGI) Annotation Pipeline, which produced the final structural and functional annotation for 18,290 predicted protein-coding genes (26). A Benchmarking Universal Single-Copy Orthologs (BUSCO) v3.0.2 (27) analysis was used to evaluate the completeness of the assembled genome based on the Stramenopile database with the Augustus (28) training set (29). The percentage of identified complete BUSCOs was 90% (100 total BUSCO groups searched; 90 complete, 8 missing). The assembly and annotation statistics are provided in Table 1. Noteworthy is that *T. minus* has a telomeric repeat sequence of TTAGGG, which differs from that of TTTAGGG reported for the species of other algal families within Xanthophyceae (30). This is the only published assembly of a yellow-green alga from the class Xanthophyceae.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number JAFCMP000000000. The version described in this paper is version JAFCMP010000000. The raw sequencing reads are deposited under the BioProject accession number PRJNA692219. The genome assembly, transcriptome, and annotations are also available from the JGI algal genome portal PhycoCosm (31) at https://phycocosm.jgi.doe.gov/Tribonema_minus/.

ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Energy (DOE) EERE/BETO via contract DE-EE0007691 (Algae Biomass Yield 2) to MicroBio Engineering, Inc., with a subcontract to Cal Poly.

The work conducted by the DOE Joint Genome Institute (JGI), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. DOE under contract number DE-AC02-05CH11231.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. DOE's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. DOE or the U.S. government.

We thank Shawn Starkenburg and Yuliya Kunde at Los Alamos National Lab for their helpful suggestions.

REFERENCES

1. Zuccarello GC, Lokhorst GM. 2005. Molecular phylogeny of the genus *Tribonema* (*Xanthophyceae*) using rbcL gene sequence data: monophyly of morphologically simple algal species. Phycologia 44:384–392. https://doi.org/10.2216/0031-8884(2005)44[384:MPOTGT]2.0.CO;2.
2. Wang H, Gao L, Chen L, Guo F, Liu T. 2013. Integration process of biodiesel production from filamentous oleaginous microalgae *Tribonema minus*. Bioresour Technol 142:39–44. https://doi.org/10.1016/j.biortech.2013.05.058.
3. Jimel M, Kviderova J, Elster J. 27 November 2020. Annual cycle of mat-forming filamentous alga *Tribonema* cf. minus (Stramenopiles, *Xanthophyceae*) in hydro-terrestrial habitats in the high Arctic revealed by multiparameter fluorescent staining. J Phycol https://doi.org/10.1111/jpy.13109.
4. Wang F, Chen J, Zhang C, Gao B. 2020. Resourceful treatment of cane sugar industry wastewater by *Tribonema minus* towards the production of valuable biomass. Bioresour Technol 316:123902. https://doi.org/10.1016/j.biortech.2020.123902.
5. Zhang Y, Wang H, Yang R, Wang L, Yang G, Liu T. 2020. Genetic

transformation of *Tribonema minus*, a eukaryotic filamentous oleaginous yellow-green alga. Int J Mol Sci 21:2106. https://doi.org/10.3390/ijms21062106.
6. Zhou W, Wang H, Zheng L, Cheng W, Gao L, Liu T. 2019. Comparison of lipid and palmitoleic acid induction of *Tribonema minus* under heterotrophic and phototrophic regimes by using high-density fermented seeds. Int J Mol Sci 20:4356. https://doi.org/10.3390/ijms20184356.
7. Wang F, Gao B, Su M, Dai C, Huang L, Zhang C. 2019. Integrated biorefinery strategy for tofu wastewater biotransformation and biomass valorization with the filamentous microalga *Tribonema minus*. Bioresour Technol 292:121938. https://doi.org/10.1016/j.biortech.2019.121938.
8. Wang H, Zhang Y, Zhou W, Noppol L, Liu T. 2018. Mechanism and enhancement of lipid accumulation in filamentous oleaginous microalgae *Tribonema minus* under heterotrophic condition. Biotechnol Biofuels 11:328. https://doi.org/10.1186/s13068-018-1329-z.
9. Wang H, Gao L, Shao H, Zhou W, Liu T. 2017. Lipid accumulation and metabolic analysis based on transcriptome sequencing of filamentous oleaginous microalgae *Tribonema minus* at different growth phases. Bioprocess Biosyst Eng 40:1327–1335. https://doi.org/10.1007/s00449-017-1791-1.
10. Zhou W, Wang H, Chen L, Cheng W, Liu T. 2017. Heterotrophy of filamentous oleaginous microalgae *Tribonema minus* for potential production of lipid and palmitoleic acid. Bioresour Technol 239:250–257. https://doi.org/10.1016/j.biortech.2017.05.045.
11. Cheng T, Zhang W, Zhang W, Yuan G, Wang H, Liu T. 2017. An oleaginous filamentous microalgae *Tribonema minus* exhibits high removing potential of industrial phenol contaminants. Bioresour Technol 238:749–754. https://doi.org/10.1016/j.biortech.2017.05.040.
12. Wang H, Gao L, Zhou W, Liu T. 2016. Growth and palmitoleic acid accumulation of filamentous oleaginous microalgae *Tribonema minus* at varying temperatures and light regimes. Bioprocess Biosyst Eng 39:1589–1595. https://doi.org/10.1007/s00449-016-1633-6.
13. Huo S, Chen J, Chen X, Wang F, Xu L, Zhu F, Guo D, Li Z. 2018. Advanced treatment of the low concentration petrochemical wastewater by *Tribonema* sp. microalgae grown in the open photobioreactors coupled with the traditional anaerobic/oxic process. Bioresour Technol 270:476–481. https://doi.org/10.1016/j.biortech.2018.09.024.
14. Huo S, Chen J, Zhu F, Zou B, Chen X, Basheer S, Cui F, Qian J. 2019. Filamentous microalgae *Tribonema* sp. cultivation in the anaerobic/oxic effluents of petrochemical wastewater for evaluating the efficiency of recycling and treatment. Biochem Eng J 145:27–32. https://doi.org/10.1016/j.bej.2019.02.011.
15. Davis AK, Anderson RS, Spierling R, Leader S, Lesne C, Mahan K, Lundquist T, Benemann JR, Lane T, Polle JEW. 2021. Characterization of a novel strain of *Tribonema minus* demonstrating high biomass productivity in outdoor raceway ponds. Bioresour Technol 331:125007. https://doi.org/10.1016/j.biortech.2021.125007.
16. Cohen Z. 1986. Products from microalgae, p 421–454. *In* Richmond A (ed), Microalgal mass culture. CRC Press, Boca Raton, FL.
17. Zhang M, Zhang Y, Scheuring CF, Wu C-C, Dong JJ, Zhang H-B. 2012. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. Nat Protoc 7:467–478. https://doi.org/10.1038/nprot.2011.455.
18. Fukasawa Y, Ermini L, Wang H, Carty K, Cheung M-S. 2020. LongQC: a quality control tool for third generation sequencing long read data. G3 (Bethesda) 10:1193–1196. https://doi.org/10.1534/g3.119.400864.
19. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
20. Chen S, Huang T, Zhou Y, Han Y, Xu M, Gu J. 2017. AfterQC: automatic filtering, trimming, error removing and quality control for fastq data. BMC Bioinformatics 18:80. https://doi.org/10.1186/s12859-017-1469-3.
21. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
22. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 13033997 [q-bioGN] https://arxiv.org/abs/1303.3997.
23. McKain MR, Wilson M. 2017. Fast-Plast: rapid de novo assembly and finishing for whole chloroplast genomes. https://github.com/mrmckain/Fast-Plast.
24. Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res 45:e18. https://doi.org/10.1093/nar/gkw955.
25. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883.
26. Kuo A, Bushnell B, Grigoriev IV. 2014. Fungal genomics: sequencing and annotation. Adv Bot Res 70:1–52. https://doi.org/10.1016/B978-0-12-397940-7.00001-X.
27. Simao 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.
28. Stanke M, Morgenstern B. 2005. AUGUSTUS: a Web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33:W465–W467. https://doi.org/10.1093/nar/gki458.
29. Hanschen ER, Hovde BT, Starkenburg SR. 2020. An evaluation of methodology to determine algal genome completeness. Algal Res 51:102019. https://doi.org/10.1016/j.algal.2020.102019.
30. Fulneckova J, Sevcikova T, Fajkus J, Lukesova A, Lukes M, Vlcek C, Lang BF, Kim E, Elias M, Sykorova E. 2013. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. Genome Biol Evol 5:468–483. https://doi.org/10.1093/gbe/evt019.
31. 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.