



Genome Sequence of the Oleaginous Green Alga, *Chlorella vulgaris* UTEX 395

Michael T. Guarnieri^{1*}, Jennifer Levering², Calvin A. Henard¹, Jeffrey L. Boore³, Michael J. Betenbaugh⁴, Karsten Zengler² and Eric P. Knoshaug¹

¹ National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO, United States, ² Division of Host-Microbe Systems and Therapeutics, Department of Pediatrics, University of California, San Diego, La Jolla, CA, United States,

³ Genome Project Solutions, Inc., Hercules, CA, United States, ⁴ Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, United States

Keywords: microalgae, biofuels, lipids, genome, *Chlorella vulgaris*

INTRODUCTION

Microalgae have garnered extensive interest as renewable fuel feedstocks due to their high production potential relative to terrestrial crops, and unique cultivation capacity on non-arable lands (Wijffels and Barbosa, 2010; Davis et al., 2011). The oleaginous chlorophyte *Chlorella vulgaris* represents a promising model microalgal system and production host, due to its ability to synthesize and accumulate large quantities of fuel intermediates in the form of storage lipids (Guarnieri et al., 2011, 2012; Gerken et al., 2013; Griffiths et al., 2014; Zuñiga et al., 2016). Recent omic analyses have identified transcriptional, post-transcriptional and -translational mechanisms governing lipid accumulation in this alga (Guarnieri et al., 2011, 2013), including active protein nitrosylation (Henard et al., 2017). Here we report the draft nuclear genome and annotation of *C. vulgaris* UTEX 395.

MATERIALS AND METHODS

Cultivation and Genomic DNA Isolation

For genomic DNA isolation *C. vulgaris* UTEX 395 was grown photoautotrophically to exponential phase in Bold's Basal Media, under constant illumination (200 $\mu\text{E m}^{-2} \text{s}^{-1}$ white fluorescent light), and supplemented with 2% CO₂/air, as described previously (Guarnieri et al., 2011, 2013). Genomic DNA was extracted following the protocol adapted from Varela-Alvarez et al. (2006).

Genome Sequencing and Assembly

Sequencing was performed using Illumina HiSeq 2000 technology with 108 cycles. 171,758,456 paired-end (SIPES) reads were trimmed to an error rate of <1:100, then trimmed until no ambiguous nucleotides remain; reads shorter than 20 nucleotides were discarded, retaining 168,611,711 reads, of which 165,874,962 remained as pairs. Resultant reads were assembled using a DeBruijn method; 113 scaffolds were generated at $\geq 1,000\times$ depth of coverage, 24 of which were longer than 100 kb and 566 of which were 20–100 kb, ultimately generating a total assembly size of 37.34 Mb, with a 61.5% GC content. This represents the smallest nuclear genome size and lowest GC content reported to date for a sequenced *Chlorella* species (Supplemental Table 1).

Genome Annotation

Transcript prediction was conducted using Maker (Cantarel et al., 2008; Campbell et al., 2014). Transcripts were six-frame translated into protein sequences and functionally annotated with EC, GO and InterProScan identifiers using two approaches. First, a bidirectional BLASTp against SwissProt sequences was carried out and paralogs were identified using BLASTclust. Secondly, InterProScan and PRIAM analyses with gene and genome-specific profiles were conducted.

To facilitate refined annotation and comprehensive pathway mapping of *C. vulgaris*, a draft nuclear genome sequence was generated and integrated with previously acquired

OPEN ACCESS

Edited by:

Qiang Wang,
Institute of Hydrobiology (Chinese
Academy of Sciences), China

Reviewed by:

Jianhua Fan,
East China University of Science and
Technology, China
Yingchun Wang,
Chinese Academy of Sciences, China

*Correspondence:

Michael T. Guarnieri
michael.guarnieri@nrel.gov

Specialty section:

This article was submitted to
Bioenergy and Biofuels,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 04 October 2017

Accepted: 19 March 2018

Published: 05 April 2018

Citation:

Guarnieri MT, Levering J, Henard CA,
Boore JL, Betenbaugh MJ, Zengler K
and Knoshaug EP (2018) Genome
Sequence of the Oleaginous Green
Alga, *Chlorella vulgaris* UTEX 395.
Front. Bioeng. Biotechnol. 6:37.
doi: 10.3389/fbioe.2018.00037

de novo transcriptomic datasets (Guarnieri et al., 2011). 7,100 transcripts were predicted from the *C. vulgaris* genome, resulting in 6,056 annotated gene models. Genomic queries identified complete gene sets encoding fatty acid and triacylglyceride biosynthetic pathways. The nitrogen assimilation inventory includes genes for nitrate/nitrite transporters and reductases. The genome also encodes meiosis-associated DMC1 and Rad51 DNA recombinase homologs (Fanning et al., 2006; Broderick et al., 2010), offering a possibility that sexual mating may occur in this microalga. Genes for the synthesis of the global stress response alarmone, guanine tetraphosphate (ppGpp) (Takahashi et al., 2004; Tozawa and Nomura, 2011), were also identified. Combined, these genetic pathways will enable potential markerless strain-engineering strategies targeting lipid accumulation in the absence of stress induction, ultimately facilitating the development of robust, deployment-viable microalgae for cost-competitive biofuel production.

DIRECT LINK TO DEPOSITED DATA AND INFORMATION TO USERS

This whole-genome project has been deposited at DDBJ/EMBL/GenBank under the accession LDKB00000000. The version described in this paper is version LDKB01000000. Additional details can be found at http://www.nrel.gov/biomass/proj_microalgal_biofuels.html and <http://chlorella.genomeprojectsolutions-databases.com>.

REFERENCES

- Broderick, S., Rehmet, K., Concannon, C., and Nasheuer, H. P. (2010). Eukaryotic single-stranded DNA binding proteins: central factors in genome stability. *Subcell. Biochem.* 50, 143–163. doi: 10.1007/978-90-481-3471-7_8
- Campbell, M. S., Holt, C., Moore, B., and Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinform.* 48, 4.11.11–14.11.39. doi: 10.1002/0471250953.bi0411s48
- Cantarel, B. L., Krof, I., Robb, S. M., Parra, G., Ross, E., Moore, B., et al. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196. doi: 10.1101/gr.6743907
- Davis, R., Aden, A., and Pienkos, P. T. (2011). Techno-economic analysis of autotrophic microalgae for fuel production. *Appl. Energy* 88, 3524–3531. doi: 10.1016/j.apenergy.2011.04.018
- Fanning, E., Klimovich, V., and Nager, A. R. (2006). A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res.* 34, 4126–4137. doi: 10.1093/nar/gkl550
- Gerken, H. G., Donohoe, B., and Knoshaug, E. P. (2013). Enzymatic cell wall degradation of *Chlorella vulgaris* and other microalgae for biofuels production. *Planta* 237, 239–253. doi: 10.1007/s00425-012-1765-0
- Griffiths, M. J., van Hille, R. P., and Harrison, S. T. (2014). The effect of nitrogen limitation on lipid productivity and cell composition in *Chlorella vulgaris*. *Appl. Microbiol. Biotechnol.* 98, 2345–2356. doi: 10.1007/s00253-013-5442-4
- Guarnieri, M. T., Laruens, L. M. L., Knoshaug, E. P., Chou, Y. C., Donohoe, B. S., and Pienkos, P. T. (2012). *Complex Systems Engineering: A Case Study for an Unsequenced Microalga*. Hoboken, NJ: John Wiley & Sons, Inc.
- Guarnieri, M. T., Nag, A., Smolinski, S. L., Darzins, A., Seibert, M., and Pienkos, P. T. (2011). Examination of triacylglycerol biosynthetic pathways via *de novo* transcriptomic and proteomic analyses in an unsequenced microalga. *PLoS ONE* 6:e25851. doi: 10.1371/journal.pone.0025851
- Guarnieri, M. T., Nag, A., Yang, S., and Pienkos, P. T. (2013). Proteomic analysis of *Chlorella vulgaris*: potential targets for enhanced lipid accumulation. *J. Proteomics* 93, 245–253. doi: 10.1016/j.jprot.2013.05.025

AUTHOR CONTRIBUTIONS

The work was designed by EK, MB, KZ, and MG. MG directed wet lab analyses. JB directed genome assembly. JL and CH directed genome annotation and pathway mapping.

ACKNOWLEDGMENTS

This work was supported by National Renewable Energy Laboratory, Laboratory-Directed Research and Development (LDRD) grants 06511103 and 06511301 and by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under Award Number DE-SC-0012658 and Office of Energy Efficiency and Renewable Energy (EERE) under Agreement No. 22000. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. We kindly thank Robert Stiles for assistance in genome assembly.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2018.00037/full#supplementary-material>

- Henard, C. A., Guarnieri, M. T., and Knoshaug, E. P. (2017). The *Chlorella vulgaris* S-Nitrosoproteome under nitrogen-replete and -deplete conditions. *Front. Bioeng. Biotechnol.* 4:100. doi: 10.3389/fbioe.2016.00100
- Takahashi, K., Kasai, K., and Ochi, K. (2004). Identification of the bacterial alarmone guanosine 5'-diphosphate 3'-diphosphate (ppGpp) in plants. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4320–4324. doi: 10.1073/pnas.0308555101
- Tozawa, Y., and Nomura, Y. (2011). Signalling by the global regulatory molecule ppGpp in bacteria and chloroplasts of land plants. *Plant Biol.* 13, 699–709. doi: 10.1111/j.1438-8677.2011.00484.x
- Varela-Alvarez, E., Andreakis, N., Lago-Leston, A., Pearson, G. A., Serrao, E. A., Procaccini, G., et al. (2006). Genomic DNA isolation from green and brown algae (Caulerpa and Fucales) for microsatellite library construction. *J. Phycol.* 42, 741–745. doi: 10.1111/j.1529-8817.2006.00218.x
- Wijffels, R. H., and Barbosa, M. J. (2010). An outlook on microalgal biofuels. *Science* 329, 796–799. doi: 10.1126/science.1189003
- Zuñiga, C., Li, C. T., Huelsman, T., Levering, J., Zielinski, D. C., McConnell, B. O., et al. (2016). Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions. *Plant Phys.* 172, 589–602. doi: 10.1104/pp.16.00593

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Guarnieri, Levering, Henard, Boore, Betenbaugh, Zengler and Knoshaug. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.