Metagenome-Assembled Genome Sequence of *Rhodobacteraceae* Bacterium Strain Clear-D3, Assembled from a *Dolichospermum* Consortium from Clear Lake, California

Chuankai Cheng, Kyra M. Florea, Eric A. Webb, J. Cameron Thrash

Department of Biological Sciences, University of Southern California, Los Angeles, California, USA

ABSTRACT

Here, we report the metagenome-assembled genome sequence of a *Rhodobacteraceae* bacterium strain, Clear-D3, that was reconstructed from a cyanobacterial enrichment from a eutrophic lake. The draft genome sequence shows evidence of an anoxygenic photoautotrophic lifestyle. Other potential capabilities include aerobic heterotrophy, flagellar motility, chemotaxis, and utilization of complex C-P compounds.

Cyanobacterial harmful algal blooms (cyanoHABs) pose an environmental threat to freshwater bodies globally. *Dolichospermum* is a common genus of cyanoHAB-forming cyanobacteria. The frequency of *Dolichospermum* blooms has increased in recent years, prompting a need to investigate and understand the biological underpinnings of bloom formation (1). We generated enrichments from *Dolichospermum* aggregates to investigate microbial interactions supporting these common harmful algal bloom (HAB) communities. Metagenomic sequencing and assembly yielded a novel *Rhodobacteraceae* metagenome-assembled genome (MAG) that will help downstream efforts to understand *Dolichospermum* symbioses.

We collected *Dolichospermum* aggregates from surface water using a bucket tow in Clear Lake, CA (lat 38.973166, long 122.72809), in August 2019. Colonies were hand separated, enriched for a total of 7 months in 50% BG-11 medium (50% BG-11 medium contains 1 part BG-11 medium and 1 part sterile Milli-Q water), and incubated at 25°C at 100 μmol Q/m²/s on a 12:12-h light/dark cycle, with NaNO₃ excluded to promote the growth of diazotrophic taxa and their symbionts. Additional medium was added to the enrichments every 2 weeks to maintain growth. Prior to sequencing, we identified the core strain of the aggregates morphologically as a *Dolichospermum* sp. (2) with an Axiostar epifluorescence microscope (Zeiss, Oberkochen, Germany). We extracted DNA from a single enrichment following an adapted Qiagen DNeasy PowerBiofilm kit protocol, modified from the manufacturer’s instructions as follows: after the addition of solution C1, samples were subjected to freeze-thaw (liquid nitrogen [−196°C] to 75°C) four times and then incubated overnight in a proteinase K solution (25 μl of a 20-mg/ml solution). Isolated DNA was verified with Tris-borate-EDTA (TBE) gel electrophoresis and quantified with NanoDrop UV-visible (UV-Vis) spectroscopy and Qubit spectrophotometry (Thermo Fisher Scientific, Waltham, MA). Illumina library preparation and paired-end (PE) 100-bp sequencing (1 Gbp) were performed with a NEBNext DNA library preparation kit according to the manufacturer’s recommendations by Novogene (Nanjing, China) with 300-bp inserts, producing 19,844,532 reads. We processed raw reads through the narrative for genome extraction from shotgun metagenome sequence data on the KBase server (3). Default settings were used for all software unless otherwise noted. We used FastQC v0.11.5 (4) to check the quality of paired-end reads, trimmed sequences with Trimmomatic v0.36 (5) to remove reads under 36 bp, assembled reads with metaSPAdes v3.13.0 (6), and binned the assembled contigs with MaxBin v2.2.4 (7). We annotated Clear-D3 through PGAP (8) and used...
GTDB-tk v1.1.0 (9) for taxonomic identification with “classify_wf” and the release 95 database. The metabolic potential was analyzed using MetaSanity v1.2 (10) and KEGG Decoder v1.1 (11).

The genome assembly pipeline yielded a 4,519,349-bp MAG we designated Clear-D3, which was taxonomically placed within the provisional genus TH137 in the family *Rhodobacteraceae*. Clear-D3 included 93 contigs (N_{50} value, 168,577 bp) with a GC content of 65.25% and 4,396 coding sequences. The average coverage of the contigs was 41.28×. CheckM v1.0.18 (12) predicted the genome as 98.44% complete with 1.73% contamination. Strain Clear-D3 has complete central carbon metabolism (glycolysis, the tricarboxylic acid cycle, and NADH-quinone oxidoreductase), showing its potential for a heterotrophic lifestyle. However, it also has genes for the Calvin-Benson-Bassham cycle, anoxygenic type II reaction centers, and sulfite dehydrogenase (quinone), suggesting the potential for photolithoautotrophy. Furthermore, this MAG shows potential in both transporting (phnCED) and utilizing (C-P lyase tetradimer complex phnJGHI) phosphonate compounds.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number JACVZW000000000. The version described in this paper is version JACVZW010000000. The BioProject number is PRJNA657201, and the reads are available at the SRA under accession number SRX8961729.

ACKNOWLEDGMENTS

This work was funded by the University of Southern California and was part of the laboratory component of BISC419, Environmental Microbiology.

We thank Elaina Graham, Ben Tully, and John F. Heidelberg for assistance with the data analysis.

C.C., K.M.F., E.A.W., and J.C.T. wrote the paper, and K.M.F. and E.A.W. are the sources of the cultures.

REFERENCES

1. Li X, Dreher TW, Li R. 2016. An overview of diversity, occurrence, genetics and toxin production of bloom-forming *Dolichospermum* (*Anabaena*) species. Harmful Algae 54:54–68. https://doi.org/10.1016/j.hal.2015.10.015.
2. Komárek J, Zapomilová E. 2008. Planktic morphospecies of the cyanobacteria *Synechococcus* (sensu lato) and *Synechocystis* sensu lato. Harmful Algae 7:324–337. https://doi.org/10.1016/j.hal.2008.09.006.
3. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Yu D. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163.
4. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Cambridge, United Kingdom.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
6. Nurk S, Meleshko D, Korobeinikov A, Pevzner PA. 2017. MetaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.
7. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605–607. https://doi.org/10.1093/bioinformatics/btv638.
8. Tatusova T, Dicuccio M, Badreddin A, Chevertvin N, Navrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nuclelic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
9. Chaumeil P-A, Mussig AJ, Henugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
10. Neely CJ, Graham ED, Tully BJ. 2020. MetaSanity: an integrated microbial genome evaluation and annotation pipeline. Bioinformatics 36:4341–4344. https://doi.org/10.1093/bioinformatics/btaa512.
11. Graham ED, Heidelberg JF, Tully BJ. 2018. Potential for primary productivity in a globally-distributed bacterial phototroph. ISME J 12:1861–1866. https://doi.org/10.1038/s41396-018-0091-3.
12. Parks DH, Imelfort M, Skennerton CT, Henugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.