AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

# Metagenome-Assembled Genome Sequence of *Vulcanococcus* sp. Strain Clear-D1, Assembled from a Cyanobacterial Enrichment Culture

**Alexis E. Floback,[a] Kyra M. Florea,[a] J. Cameron Thrash,[a] Eric A. Webb[a]**

[a]Department of Biological Sciences, University of Southern California, Los Angeles, California, USA

**ABSTRACT** We report the metagenome-assembled genome sequence of a *Vulcanococcus* sp. binned from a cyanobacterial enrichment culture. The genome contains 39 contigs comprising 2.96 Mbp and is estimated as 100% complete, with a GC content of 63.9% and 3,261 predicted coding genes.

Formation of harmful algal blooms (HABs) at the volcanically active Clear Lake in California has been reported for decades (1); the HABs are dominated by diazotrophic cyanobacteria, such as *Aphanizomenon* and *Dolichospermum* (2), which are known toxin producers (3, 4). Clear Lake is a source of drinking water for local communities and brings in over 50 million dollars annually through recreational activities and tourism (5). To better understand microbial interactions supporting *Dolichospermum* communities, we performed enrichment culturing and metagenomic sequencing, assembly, and binning. Here, we report the metagenome-assembled genome (MAG) for a species of *Vulcanococcus*, a genus that was only recently isolated from volcanic Lake Albano in Italy (6).

A *Dolichospermum* enrichment was collected at Clear Lake (lat 38.973167, long 122.728089), by a surface bucket tow in August 2019. Hand-picked *Dolichospermum* colonies were cultured for 7 months in 50% BG-$11_0$ medium (7)/50% sterile Milli-Q water and incubated at 25°C (100 $\mu$mol Q/m$^2$/s) on a 12:12-h light/dark cycle, with no $NaNO_3$ added to enrich for diazotrophs. Additional medium was added approximately every 2 weeks to maintain growth. Genomic DNA was extracted from the enrichment community with the DNeasy PowerBiofilm kit (Qiagen) following the manufacturer's instructions but with the addition of five freeze-thaw cycles and a subsequent overnight incubation at 55°C with 25 $\mu$l of 20 mg/ml proteinase K and solution C1 from the kit. Isolated DNA was verified with Tris-borate-EDTA (TBE) gel electrophoresis and quantified with NanoDrop UV-visible (UV-Vis) spectroscopy and Qubit spectrofluorometry (Thermo Fisher Scientific, Waltham, MA). Illumina paired-end (PE) 150-bp sequencing (1 Gbp) was performed by Novogene using 300-bp inserts, after library preparation with a NEBNext DNA library preparation kit according to the manufacturer's recommendations. This resulted in 19,844,532 reads. KBase (8) and modules within were used for assembly, as follows. The quality of paired-end reads was checked with FastQC v0.11.5 (9), and sequences were trimmed with Trimmomatic v0.36 (10) with reads under 36 bp being removed. Metagenome assembly was performed with metaSPAdes v3.13.0 (11), and binning was completed with MaxBin v2.2.4 (12). Taxonomy was assigned with GTDB-tk v1.1.0, run with the parameter "classify_wf" and using the release 95 database (13). Default settings were used for all software unless otherwise noted.

One metagenomic bin (Clear-D1) from the enrichment comprised 2,960,550 bp (GC content, 63.9%) in 39 contigs with an $N_{50}$ value of 149,920 bp. CheckM v1.0.18 (14) estimated Bin001 as 100% complete with 0.54% contamination, and GTDB-tk classified it as a *Vulcanococcus* sp. The genome was annotated with PGAP v4.11 (15), which

**Citation** Floback AE, Florea KM, Thrash JC, Webb EA. 2020. Metagenome-assembled genome sequence of *Vulcanococcus* sp. strain Clear-D1, assembled from a cyanobacterial enrichment culture. Microbiol Resour Announc 9:e01121-20. https://doi.org/10.1128/MRA.01121-20.

**Editor** Frank J. Stewart, Georgia Institute of Technology



Address correspondence to Eric A. Webb, eawebb@usc.edu.

**Received** 24 September 2020
**Accepted** 25 September 2020
**Published** 3 December 2020

predicted 3,083 coding genes, 57 pseudogenes, and 45 noncoding RNA sequences. MetaSanity v3.0 (16) analysis using the FuncSanity module revealed that the *Vulcanococcus* sp. strain Clear-D1 genome has all of the genes required for photoautotrophy but also has genes encoding proteins for sulfide oxidation, sulfur assimilation, and arsenic reduction. Nitrogen fixation genes were missing, unlike in another isolate of this genus (6), and the genome contained predicted genes for thiamine, riboflavin, cobalamin, and retinal biosynthesis. The genome also contained genes for putative ferrous iron transporters and genes for proteins with ferric iron ABC-type substrate-binding capabilities. Thus, this species appears to have competitive nutrient acquisition strategies and interesting capabilities for secondary metabolism that reflect the volcanic activity at Clear Lake.

**Data availability.** This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number JACVZV000000000. The version described in this paper is version JACVZV010000000. 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.

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

## REFERENCES

1. Horne AJ, Javornicky P, Goldman CR. 1971. A freshwater "red tide" on Clear Lake, California 1. Limnol Oceanogr 16:684–689. https://doi.org/10.4319/lo.1971.16.4.0684.
2. Kurobe T, Baxa DV, Mioni CE, Kudela RM, Smythe TR, Waller S, Chapman AD, Teh SJ. 2013. Identification of harmful cyanobacteria in the Sacramento-San Joaquin Delta and Clear Lake, California by DNA barcoding. Springerplus 2:491. https://doi.org/10.1186/2193-1801-2-491.
3. 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.
4. Cirés S, Ballot A. 2016. A review of the phylogeny, ecology and toxin production of bloom-forming *Aphanizomenon* spp. and related species within the *Nostocales* (cyanobacteria). Harmful Algae 54:21–43. https://doi.org/10.1016/j.hal.2015.09.007.
5. Goldstein JJ, Tolsdorf TN. 1994. An economic analysis of potential water quality improvement in Clear Lake: benefits and costs of sediment control, including a geological assessment of potential sediment control levels: Clear Lake Basin, Lake County, California. U.S. Department of Agriculture, Washington, DC.
6. Di Cesare A, Cabello-Yeves PJ, Chrismas NA, Sánchez-Baracaldo P, Salcher MM, Callieri C. 2018. Genome analysis of the freshwater planktonic *Vulcanococcus limneticus* sp. nov. reveals horizontal transfer of nitrogenase operon and alternative pathways of nitrogen utilization. BMC Genomics 19:259. https://doi.org/10.1186/s12864-018-4648-3.
7. Andersen RA (ed). 2005. Algal culturing techniques. Elsevier, Burlington, MA.
8. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo 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.
9. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Cambridge, United Kingdom.
10. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
11. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834. https://doi.org/10.1101/gr.213959.116.
12. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2:26. https://doi.org/10.1186/2049-2618-2-26.
13. Chaumeil P-A, Mussig AJ, Hugenholtz 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.
14. Parks DH, Imelfort M, Skennerton CT, Hugenholtz 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.
15. Tatusova T, Dicuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
16. 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.