

Draft Genome Sequence of an Oscillatorian Cyanobacterium, Strain ESFC-1

R. Craig Everrood,^a Dagmar Wobken,^{a,b*} Steven W. Singer,^c Luke C. Burow,^{a,b} Nikos Kyrpides,^d Tanja Woyke,^d Lynne Goodwin,^e Angela Detweiler,^a Leslie Prufert-Bebout,^a Jennifer Pett-Ridge^f

Exobiology Branch, NASA Ames Research Center, Moffett Field, California, USA^a; Departments of Chemical Engineering and Civil and Environmental Engineering, Stanford University, Stanford, California, USA^b; Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA^c; DOE Joint Genome Institute, Walnut Creek, California, USA^d; Los Alamos National Laboratory, Los Alamos, New Mexico, USA^e; Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, California, USA^f

* Present address: Dagmar Wobken, Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, University of Vienna, Vienna, Austria.

R.C.E. and D.W. contributed equally to this work.

The nonheterocystous filamentous cyanobacterium strain ESFC-1 has recently been isolated from a marine microbial mat system, where it was identified as belonging to a recently discovered lineage of active nitrogen-fixing microorganisms. Here, we report the draft genome sequence of this isolate. The assembly consists of 3 scaffolds and contains 5,632,035 bp with a GC content of 46.5%.

Received 13 June 2013 Accepted 24 June 2013 Published 1 August 2013

Citation Everrood RC, Wobken D, Singer SW, Burow LC, Kyrpides N, Woyke T, Goodwin L, Detweiler A, Prufert-Bebout L, Pett-Ridge J. 2013. Draft genome sequence of an oscillatorian cyanobacterium, strain ESFC-1. *Genome Announc.* 1(4):e00527-13. doi:10.1128/genomeA.00527-13.

Copyright © 2013 Everrood et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Jennifer Pett-Ridge, pettridge2@llnl.gov.

Nitrogen-fixing filamentous cyanobacteria are central components of intertidal microbial mat communities (1, 2). Recently, a new and diverse lineage of filamentous nitrogen-fixing cyanobacteria lacking both heterocysts and an extracellular sheath was identified as the major cyanobacterial diazotroph in the intertidal microbial mats at Elkhorn Slough, Moss Landing, California (3). An isolate of this lineage, strain ESFC-1, shares only a low 16S rRNA gene identity (<95%) with other identified cyanobacteria.

Strain ESFC-1 was isolated from the upper 2 mm of mat samples originating from Elkhorn Slough, California (36°48'46.61''N, 121°47'4.89''W). For isolation, mat material was plated onto nitrogen-free ASN and modified ASN agar plates and was subsequently rendered pure in liquid ASN (3, 4). High-molecular-weight genomic DNA was isolated through lysis with lysozyme, proteinase K, and SDS based on the protocol for bacterial genomic DNA isolation using cetyltrimethylammonium bromide (CTAB) provided by the Joint Genome Institute (JGI) (<http://my.jgi.doe.gov/general/protocols/JGI-Bacterial-DNA-isolation-CTAB-Protocol-2012.pdf>). RNA was digested with RNase according to this protocol, and 50 µg of DNA was provided for sequencing. The draft genome sequence of strain ESFC-1 was generated by the U.S. Department of Energy (DOE) JGI using Illumina sequencing technology (5). General aspects of library construction and sequencing performed by JGI are available at <http://www.jgi.doe.gov/>. Both an Illumina short-insert paired-end library with an average insert size of 222 bp, which generated 15,283,374 reads, and an Illumina long-insert paired-end library with an average insert size of 7,791 bp, which generated 18,062,354 reads, were constructed and sequenced, totaling 4,099 Mbp of Illumina data (6).

The initial draft data were assembled with Allpaths, version

r38445 (7), and the consensus was computationally shredded into 10-kbp overlapping fake reads (shreds). The Illumina draft data were also assembled with Velvet, version 1.1.05 (8), and the consensus sequences were computationally shredded into 1.5-kbp overlapping fake reads (shreds). The draft data were assembled again with Velvet using the shreds from the first assembly to guide the next assembly. The consensus from the second Velvet assembly was shredded into 1.5-kbp overlapping fake reads. Fake reads from the Allpaths assembly, both Velvet assemblies, and a subset of the Illumina CLIP paired-end reads were assembled using parallel Phrap version 4.24 (High Performance Software, LLC). Possible misassemblies were checked and manually corrected in Consed (9, 10, 11). The final assembly is based on 4,099 Mbp of Illumina draft data, with an average of 719× coverage of the genome. ESFC-1 was resolved to 3 scaffolds consisting of 5,431,811, 135,349, and 64,875 bp (5,632,035 bp total). Average GC content was 46.51%.

Automated annotation was performed with both the RAST annotation server and the Integrated Microbial Genomes (IMG) system (12, 13). IMG identified 4,914 candidate protein-encoding genes, of which 71.27% had a predicted function. The ESFC-1 genome contains 72 tRNA genes and 2 rRNA operons. These rRNA operons have an average GC content of 57%. As predicted by RAST, these operons (1,451 and 1,452 bp) are 98.5% similar. The closest sequence match to either small-subunit (SSU) gene is from the marine unicellular *Aphanocapsa* sp. HBC6 at 93.6% similarity (accession EU249123 [14]).

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession number [ARCP00000000](http://www.ncbi.nlm.nih.gov/nuccore/ARCP00000000). The version described in this paper is the first version.

ACKNOWLEDGMENTS

We extend our sincere thanks to Brad Bebout and Alfred Spormann for mentorship and hosting of much of the lab work. We also thank Jeff Cann, Associate Wildlife Biologist, Central Region, California Department of Fish and Game, for coordinating our access to the Moss Landing Wildlife Area.

Funding was provided by the U.S. Department of Energy (DOE) Genomic Science Program under contract SCW1039. D.W. was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft). Sequencing and support were provided by scientists at the DOE Joint Genome Institute. Work conducted by the U.S. Department of Energy Joint Genome Institute was supported by the Office of Science of the U.S. Department of Energy under contract number DE-AC02-05CH11231. Work at LLNL was performed under the auspices of the U.S. Department of Energy at Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Work at LBNL was performed under the auspices of the U.S. Department of Energy at Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231.

REFERENCES

1. Bebout BM, Paerl HW, Crocker KM, Prufert LE. 1987. Diel interactions of oxygenic photosynthesis and N₂ fixation (acetylene reduction) in a marine microbial mat community. *Appl. Environ. Microbiol.* **53**: 2353–2362.
2. Stal LJ. 1995. Physiological ecology of cyanobacteria in microbial mats and other communities. *New Phytol.* **131**:1–32.
3. Woebken D, Burow LC, Prufert-Bebout L, Bebout BM, Hoehler TM, Pett-Ridge J, Spormann AM, Weber PK, Singer SW. 2012. Identification of a novel cyanobacterial group as active diazotrophs in a coastal microbial mat using NanoSIMS analysis. *ISME J.* **6**:1427–1439.
4. Rippka R. 1988. Isolation and purification of cyanobacteria. *Methods Enzymol.* **167**:3–27.
5. Bennett S. 2004. Solexa Ltd. *Pharmacogenomics* **5**:433–438.
6. Zhang X, Davenport KW, Gu W, Daligault HE, Munk AC, Tashima H, Reitenga K, Green LD, Han CS. 2012. Improving genome assemblies by sequencing PCR products with PacBio. *BioTechniques* **53**:61–62.
7. Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB. 2008. ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.* **18**:810–820.
8. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**:821–829.
9. Ewing B, Green P. 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res.* **8**:186–194.
10. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Res.* **8**:175–185.
11. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
12. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* **9**:75.
13. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I, Lykidis A, Mavromatis K, Ivanova N, Kyrpides NC. 2006. The integrated microbial genomes (IMG) system. *Nucleic Acids Res.* **34**:344–348.
14. Foster JS, Green SJ, Ahrendt SR, Golubic S, Reid RP, Hetherington KL, Bebout L. 2009. Molecular and morphological characterization of cyanobacterial diversity in the stromatolites of Highborne Cay, Bahamas. *ISME J.* **3**:573–587.