Draft Genome Sequence of the Filamentous Cyanobacterium *Leptolyngbya* sp. Strain Heron Island J, Exhibiting Chromatic Acclimation

Robin Paul,[a] Robert E. Jinkerson,[b] Kristina Buss,[c] Jason Steel,[c] Remus Mohr,[d] Wolfgang R. Hess,[d] Min Chen,[e] Petra Fromme[a]

Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona, USA[a]; Department of Chemistry and Geochemistry, Colorado School of Mines, Golden, Colorado, USA[b]; Biodesign Institute, Arizona State University, Tempe, Arizona, USA[c]; Genetics and Experimental Bioinformatics, University of Freiburg, Institute of Biology III, Freiburg, Germany[d]; School of Biological Sciences, the University of Sydney, Sydney, NSW, Australia[e]

***Leptolyngbya* sp. strain Heron Island is a cyanobacterium exhibiting chromatic acclimation. However, this strain has strong interactions with other bacteria, making it impossible to obtain axenic cultures for sequencing. A protocol involving an analysis of tetranucleotide frequencies, G+C content, and BLAST searches has been described for separating the cyanobacterial scaffolds from those of its cooccurring bacteria.**

Address correspondence to Petra Fromme, pfromme@asu.edu.

*L*eptolyngbya sp. strain Heron Island is a cyanobacterium classified in section III, a section that consists of filamentous cyanobacteria that grow in a single plane (1). This cyanobacterium was isolated from Heron Island, Australia, in 2009. *Leptolyngbya* sp. Heron Island exhibits chromatic acclimation, a phenomenon in which the composition of the light-harvesting complex phycobilisome (PBS) is modified with changes in light conditions (2–4).

DNA was isolated from *Leptolyngbya* sp. Heron Island using the Qiagen DNeasy plant minikit. The genomic DNA was broken into fragments by sonication for 5 min. The genome of *Leptolyngbya* sp. Heron Island was sequenced using an Illumina HiSeq 2000, applying the short-insert paired-end sequencing protocol (100-bp insert size) to a depth of coverage of 100×. These reads were *de novo* assembled using the ABySS (version 1.3.6) software (5). *Leptolyngbya* sp. Heron Island cultures cannot be grown axenically; therefore, the genomic assembly harbors sequences from other heterotrophic bacteria that are associated with this cyanobacterium. Several approaches were used to detect and separate any heterotrophic bacterial sequences or chimeras from the *Leptolyngbya* sp. Heron Island genome assembly. They include a G+C percentage analysis, a tetranucleotide frequency analysis, comparison to a reference cyanobacterial genome, and an analysis of gene annotations. The initial G+C percentage analysis of the assembled scaffolds indicated that 3 to 4 contaminating genomes are likely present. The first step consisted of devising a BLAST algorithm that selected only those scaffolds that contained a gene that matched some gene in *Leptolyngbya* sp. strain PCC 7376. The tetranucleotide frequencies (6) of the BLAST positive scaffolds were calculated using TETRA (7), followed by principal component analysis (8) using Biopython (9) of the resultant matrix obtained from TETRA. Finally, the scaffolds were separated on the basis of their G+C percentage to yield scaffolds that contained only cyanobacterial genes.

Genome annotation was carried out using GeneMark S (10)

and the NCBI Prokaryotic Genomes Annotation Pipeline (PGAP) 2.0. The resulting gene models were validated against the NCBI nonredundant protein database by BLASTp, which yielded maximum sequence homology with cyanobacterial genes.

The final assembly of *Leptolyngbya* sp. Heron Island contains 119 scaffolds, with an N_{50} scaffold length of 103,122 bp and an overall assembly size of 8.06 Mb, making it one of the biggest genome sequences among cyanobacteria (11). The annotation includes 7,223 estimated protein-coding sequences, 55 tRNAs, and 9 clustered regularly interspaced short palindromic repeat (CRISPR) arrays. The latter component suggests an amazing capability to deal with attacking bacteriophages, as CRISPR arrays are a key defense mechanism against invading DNA (12), and it is consistent with a report that subsection III and IV cyanobacteria tend to have a higher number of CRISPR loci and of repeat-spacer units (13). The availability of the sequences of the phycobilisome genes in *Leptolyngbya* sp. Heron Island will help in the refinement of phycoerythrin and phycocyanin crystal structure obtained from this organism. Eventually, with the detailed phycobilisome-encoding genes, the genome information will also help in studying chromatic acclimation in this newly isolated cyanobacterium.

Nucleotide sequence accession number. The draft genome sequence of *Leptolyngbya* sp. strain Heron Island J has been deposited at DDBJ/EMBL/GenBank under the accession no. AWNH00000000.1. The version described in this paper is the first version.

ACKNOWLEDGMENTS

REFERENCES

1. **Schirrmeister BE, Antonelli A, Bagheri HC.** 2011. The origin of multi-cellularity in cyanobacteria. BMC Evol. Biol. **11:**45. http://dx.doi.org/10.1186/1471-2148-11-45.

2. **Kehoe DM, Gutu A.** 2006. Responding to color: the regulation of complementary chromatic adaptation. Annu. Rev. Plant Biol. **57:**127–150. http://dx.doi.org/10.1146/annurev.arplant.57.032905.105215.

3. **Tandeau de Marsac N.** 2003. Phycobiliproteins and phycobilisomes: the early observations. Photosynth. Res. **76:**193–205. http://dx.doi.org/10.1023/A:1024954911473.

4. **Gutu A, Kehoe DM.** 2012. Emerging perspectives on the mechanisms, regulation, and distribution of light color acclimation in cyanobacteria. Mol. Plant **5:**1–13. http://dx.doi.org/10.1093/mp/ssr054.

5. **Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I.** 2009. ABySS: a parallel assembler for short read sequence data. Genome Res. **19:**1117–1123. http://dx.doi.org/10.1101/gr.089532.108.

6. **Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO.** 2004. Application of tetranucleotide frequencies for the assignment of genomic fragments. Environ. Microbiol. **6:**938–947. http://dx.doi.org/10.1111/j.1462-2920.2004.00624.x.

7. **Teeling H, Waldmann J, Lombardot T, Bauer M, Glöckner FO.** 2004. Tetra: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences. BMC Bioinformatics **5:**163. http://dx.doi.org/10.1186/1471-2105-5-163.

8. **Ringnér M.** 2008. What is principal component analysis? Nat. Biotechnol. **26:**303–304. http://dx.doi.org/10.1038/nbt0308-303.

9. **Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ.** 2009. Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics **25:**1422–1423. http://dx.doi.org/10.1093/bioinformatics/btp163.

10. **Besemer J, Lomsadze A, Borodovsky M.** 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. **29:**2607–2618. http://dx.doi.org/10.1093/nar/29.12.2607.

11. **Shih PM, Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA.** 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc. Natl. Acad. Sci. U. S. A. **110:**1053–1058. http://dx.doi.org/10.1073/pnas.1217107110.

12. **Sorek R, Kunin V, Hugenholtz P.** 2008. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat. Rev. Microbiol. **6:**181–186. http://dx.doi.org/10.1038/nrmicro1793.

13. **Cai F, Axen SD, Kerfeld CA.** 2013. Evidence for the widespread distribution of CRISPR-Cas system in the phylum *Cyanobacteria*. RNA Biol. **10:**687–693. http://dx.doi.org/10.4161/rna.24571.