Finished Genome Sequence of the Unicellular Cyanobacterium *Synechocystis* sp. Strain PCC 6714

Matthias Kopf,[a] Stephan Klähn,[a] Björn Voss,[a] Kurt Stüber,[b] Bruno Huettel,[b] Richard Reinhardt,[b] Wolfgang R. Hess[a]

Genetics and Experimental Bioinformatics, Faculty of Biology, University of Freiburg, Freiburg, Germany[a]; Max Planck-Genome-Centre, Cologne, Germany[b]

***Synechocystis* sp. strain PCC 6714 is a unicellular cyanobacterium closely related to the popular model organism *Synechocystis* sp. strain PCC 6803. A combination of PacBio SMRT and Illumina GAIIx data results in a highly accurate finished genome sequence that provides a reliable resource for further comparative analyses.**

Address correspondence to Wolfgang R. Hess, wolfgang.hess@biologie.uni-freiburg.de.

*S*ynechocystis sp. strain PCC 6714 (from here on *Synechocystis* 6714) is a unicellular cyanobacterium and closely related to the popular model organism *Synechocystis* sp. strain PCC 6803 (*Synechocystis* 6803). Despite the large number of recently sequenced cyanobacterial genomes (1), there is still no finished genome available for comparative studies with *Synechocystis* 6803. Recent studies based on a draft genome of *Synechocystis* 6714 have already revealed the potential for comparative analysis by explaining its lower salt tolerance (2) as well as differences in two clustered regularly interspaced short palindromic repeats (CRISPR) associated proteins (CRISPR-*cas*) loci (3). *Synechocystis* 6714 has also been investigated for various other unique aspects (4–7).

For SMRT sequencing, high-quality genomic DNA was isolated using the CTAB protocol (8). Libraries were prepared according to the large SMRTbell gDNA protocol (Pacific Biosciences) with 10 kb insert size. Genomic DNA was sequenced with a PacBio RS II platform yielding 134,034 reads with an average length of 3,818 nt. The reads were *de novo* assembled with HS HGAP Assembly version 2 (Pacific Biosciences), resulting in 7 contigs with a 97-fold average coverage. Three of the contigs could be identified as assembly artifacts and were removed. In order to maximize sequence quality, the publicly available Illumina GAIIx-based draft sequences of *Synechocystis* 6714 (2) were fragmented into pieces of 30 nt and mapped against the SMRT-based assembly using the short read mapper segemehl (version 0.1.7-403) (9). All mismatches, insertions, and deletions were corrected in favor of the highly accurate draft assembly sequences. The final contigs were checked for circularization and overlapping ends were trimmed. Gene prediction and annotation were done with RAST (10).

The finished genome of *Synechocystis* 6714 consists of 4 circular contigs. The largest contig represents the chromosome (3.5 Mb) while the other three contigs represent the plasmids pSYLA (109 kb), pSYLB (104 kb), and pSYLC (41 kb). The annotation revealed 3,770 protein-coding sequences, 42 tRNAs, two rRNA clusters, and three loci of CRISPR-*cas* genes located on the plasmids pSYLA and pSYLB. The finding of the putative prophage Psy1, as well as the absence of the *ggtABCD* and *pilA5* genes in the former draft assembly (2), was confirmed in the finished assembly. Additionally, 37 new coding sequences (CDSs), one tRNA, and a second rRNA cluster were detected, compared to the draft genome. The availability of the highly accurate finished genome sequence of *Synechocystis* 6714 provides a reliable resource for future comparative analysis with *Synechocystis* 6803, the most advanced cyanobacterial model.

Nucleotide sequence accession numbers. The finished genome sequences have been deposited at DDBJ/ENA/GenBank under the accession no. CP007542 (chromosome), CP007543 (pSYLA), CP007544 (pSYLB), and CP007545 (pSYLC). The versions described in this paper are the first versions.

REFERENCES

1. 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. USA **110:**1053–1058. http://dx.doi.org/10.1073/pnas.1217107110.
2. Kopf M, Klähn S, Pade N, Weingärtner C, Hagemann M, Voß B, Hess WR. 2014. Comparative genome analysis of the closely related *Synechocystis* strains PCC 6714 and PCC 6803. DNA Res. **21:**255–266. http://dx.doi.org/10.1093/dnares/dst055.
3. Hein S, Scholz I, Voß B, Hess WR. 2013. Adaptation and modification of three CRISPR loci in two closely related Cyanobacteria. RNA Biol. **10:**852–864. http://dx.doi.org/10.4161/rna.24160.
4. Ajlani G, Kirilovsky D, Picaud M, Astier C. 1989. Molecular analysis of psbA mutations responsible for various herbicide resistance phenotypes in Synechocystis 6714. Plant Mol. Biol. **13:**469–479. http://dx.doi.org/10.1007/BF00027307.
5. Vernotte C, Picaud M, Kirilovsky D, Olive J, Ajlani G, Astier C. 1992. Changes in the photosynthetic apparatus in the cyanobacterium *Synechocystis* sp. PCC 6714 following light-to-dark and dark-to-light transitions. Photosynth. Res. **32:**45–57. http://dx.doi.org/10.1007/BF00028797.
6. Imashimizu M, Fujiwara S, Tanigawa R, Tanaka K, Hirokawa T, Nakajima Y, Higo J, Tsuzuki M. 2003. Thymine at -5 is crucial for cpc

promoter Activity of *Synechocystis* sp. Strain PCC 6714. J. Bacteriol. **185:** 6477–6480. http://dx.doi.org/10.1128/JB.185.21.6477-6480.2003.

7. **Campanini B, Schiaretti F, Abbruzzetti S, Kessler D, Mozzarelli A.** 2006. Sulfur mobilization in cyanobacteria: the catalytic mechanism of L-cystine C-S lyase (C-DES) from *Synechocystis*. J. Biol. Chem. **281:** 38769–38780. http://dx.doi.org/10.1074/jbc.M607098200.

8. **Wilson K.** 2001. Preparation of Genomic DNA from Bacteria. *In* Current protocols in molecular Biology. John Wiley & Sons, New York, NY. http://dx.doi.org/10.1002/0471142727.mb0204s56.

9. **Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermüller J.** 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLOS Comput. Biol. **5:**e1000502. http://dx.doi.org/10.1371/journal.pcbi.1000502.

10. **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. http://dx.doi.org/10.1186/1471-2164-9-75.