Draft Genome Sequence of the Toxic Freshwater *Microcystis aeruginosa* Strain PMC 728.11 (*Cyanobacteria*, *Chroococcales*)

Sébastien Halary,[a] Charlotte Duval,[a] Alison Gallet,[a] Sébastien Duperron,[a] Bérénice Piquet,[a] Justine Demay,[a] Cécile Bernard,[a] Benjamin Marie[a]

[a]UMR 7245 MCAM MNHN-CNRS, Muséum National d'Histoire Naturelle, Paris, France

ABSTRACT *Microcystis aeruginosa* is one of the major species that cause toxic cyanobacterial blooms in freshwater systems worldwide. Here, we report the draft genome sequence of *M. aeruginosa* PMC 728.11, a microcystin-producing cyanobacterium isolated from the freshwater reservoir of Juanon in Valence, France. The genome sequence contains 276 contigs, consisting of 5,536,025 bp and 5,594 putative protein-coding genes, among which are several biosynthetic gene clusters encoding enzyme complexes involved in the production of various bioactive and toxic metabolites.

Microcystis is among the most widespread cyanobacterial genera worldwide and is frequently reported as responsible for bloom events in freshwater environments. These blooms generally occur when water temperatures exceed 15°C in water bodies enriched by anthropogenic nutrient loading (1). Because members of this genus are able to produce diverse toxic compounds, including potent hepatotoxic microcystins, *Microcystis* recurrent blooms pose a risk for populations using impaired water resources for drinking water supplies, recreational activities, and fisheries (2). Thus, *Microcystis* strains have become good models for the investigation of ecotoxicological impacts induced by *Microcystis* blooms on aquatic organisms (3, 4).

Microcystis aeruginosa strain PMC 728.11 was isolated in September 2011 from the Juanon artificial pond (44°82′99″N, 5°01′55″E; Valence, France) during an intense bloom event. Briefly, water sample was spread onto BG11 agar plates (12:12 h light/dark cycle, 20°C), and then individual colonies were picked and grown in liquid BG11 medium. The production of microcystins was detected by enzyme-linked immunosorbent assay (ELISA) (with AD4G2 antibody; Abraxis, USA) and high-resolution mass spectrometry, together with the detection of two PCR amplicons that are commonly used as markers of its biosynthesis and corresponding to *mcyA* and *mycE* genes (5). The clonal, but nonaxenic, strain was cultured in BG-11 medium (6) at 25°C in 250-ml Erlenmeyer vessels, with a photon flux density of $12\,\mu\mathrm{mol} \cdot \mathrm{m}^{-2} \cdot \mathrm{s}^{-1}$ and a 12:12-h light/dark cycle. Total DNA extraction was carried out using a ZymoBIOMICS DNA minikit (Zymo Research, CA), and sequencing was done using 2×250-bp reads from both an Illumina HiSeq 2500 instrument after an initial preparation of the library (Nextera XT sample kit) and a single-molecule real-time PacBio RS II platform after library preparation with the SMRTbell library using the Express template prep kit (Pacific Biosciences). Raw reads were inspected, cut, and filtered using FastQC v0.11.5, Cutadapt v1.15, and Prinseq v0.20.4, respectively (7–9) (resulting Illumina reads: 4,948,014 reads, N_{50} value of 235 bp, coverage of 97×; resulting PacBio SR2 reads: 86,577 reads, N_{50} value of 10,986 bp, coverage of 63×). Scaffolds were assembled from HiSeq and PacBio reads using a SPAdes-based Unicycler hybrid assembler with default parameters (10, 11). Nodes from assembly graphs were clustered using MyCC (k-mer size, 4; minimal sequence size, 1,000) and taxonomically annotated using the Contig Annotation Tool (12). 16S rRNA-encoding genes were also extracted from these nodes using Metaxa 2 and then annotated using ACT (13).
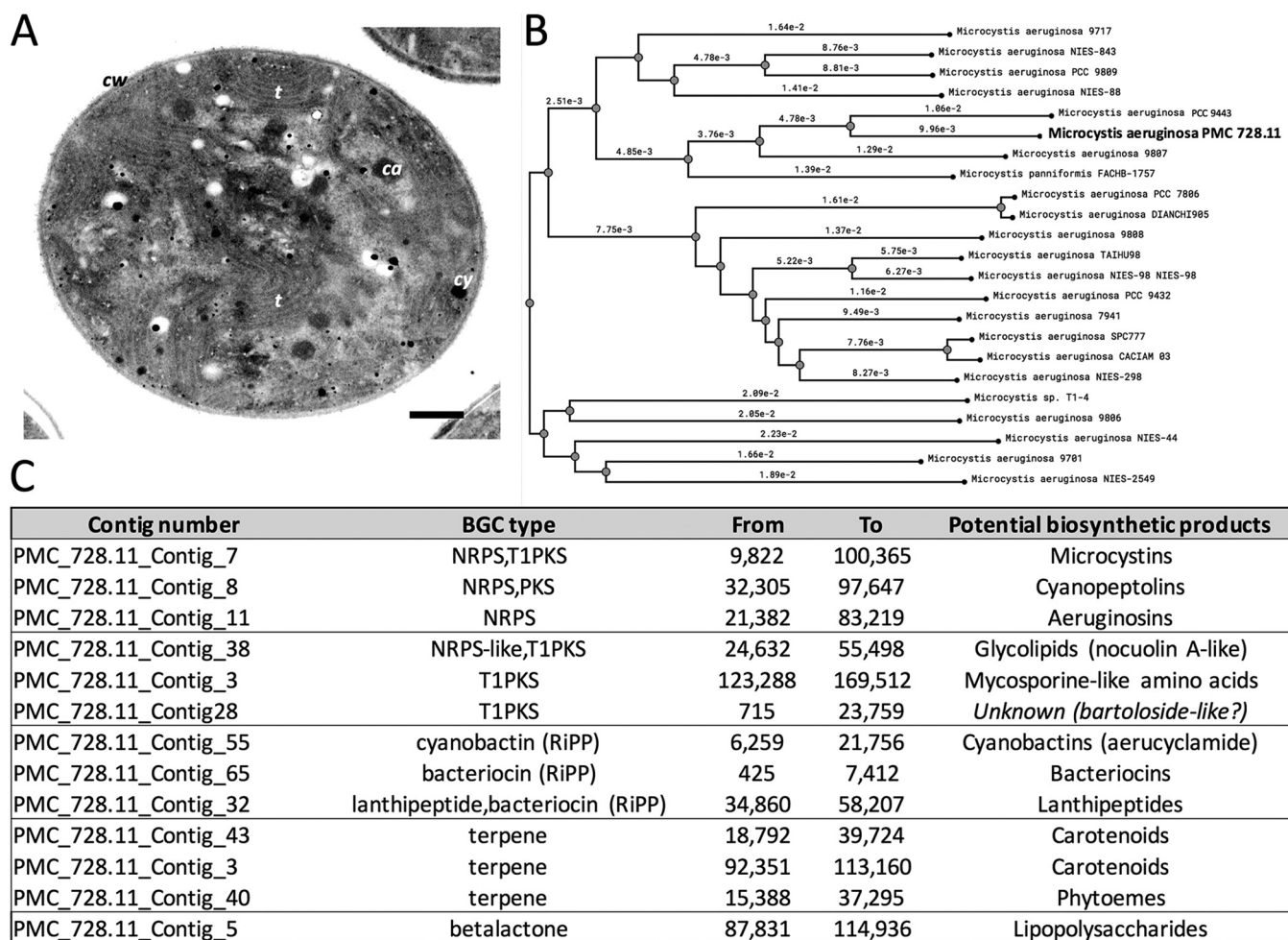
FIG 1 *Microcystis aeruginosa* strain PMC 728.11. (A) Transmission electron micrograph displaying ultrastructural details of a *Microcystis* cell, presently under division (bar, 500 nm; ca, carboxysome; cy, cyanophycine granule; cw, cell wall; t, thylakoid). (B) Phylogram of available *Microcystis* genomes based on a rapid neighbor-joining algorithm (v1.0.4) using default parameters (19) with average nucleotide identity (ANI) distances (PMC 728.11 is indicated in bold). (C) Biosynthetic gene clusters and their potential products detected by antiSMASH from a PMC 728.11 genome search, comprising gene clusters producing microcystins, among other bioactive compounds (20).

All contigs were pairwise aligned using MegaBLAST (E value ≤ 1e-10), and all sequences sharing a ≥98% similarity on the shortest sequence were considered as coming from the same genome. Congruent data between these diverse methodologies (binning with MyCC and BLAST with CAT) allowed us to characterize the draft genome sequence of *Microcystis aeruginosa* PMC 728.11.

The *Microcystis aeruginosa* PMC 728.11 genome completeness and contamination estimated from the genomes as assessed using CheckM v1.13 with default parameters (14) were 98.57% and 1.02%, respectively. Annotation was performed using the MicroScope platform (15). The PMC 728.11 genome comprised 276 contigs (maximum length, 175,236 bp; N_{50}, 6,601 bp; coding ratio, 78.3%) representing 5.536 Mbp, with a GC content of 42.4%. It contains potentially 5,594 gene features, including 40 tRNAs, a complete 16S-23S-5S rRNA operon, and 5 CRISPRs according to a CheckM search, these features being in perfect agreement with the 22 *Microcystis* genomes publicly available on the MicroScope server (2). Calculation of the average nucleotide identity (ANI) based on the BLAST algorithm (16) showed that PMC 728.11 displays 97.95% similarity to *M. aeruginosa* strain PCC 9443, collected in 1994 from a pond in Landjia, Central African Republic (Fig. 1B) (2).

Specialized metabolite biosynthetic gene clusters (BGCs) were identified using antiSMASH v5.1.2 (17) and MIBiG v1.4 (18). The genome harbors several gene clusters

involved in the biosynthesis of various cyanopeptides, including the cyanotoxin microcystin, in addition to cyanopeptolins, aeruginosins (encoded by the nonribosomal peptide synthetase/polyketide synthase [NRPS/PKS] pathways), and cyanobactins, bacteriocins, and lanthipeptides (encoded by ribosome-synthesized posttranslationally modified peptide [RiPP] pathways), together with other genes encoding enzymes involved in the biosynthesis of glycolipids, mycosporine-like amino acids, carotenoids, phytoenes, and lipopolysaccharides (2).

In conclusion, *M. aeruginosa* strain PMC 728.11 displays features typical of toxin-producing *Microcystis*. This strain represents a promising model for further investigation and characterization of bioactive metabolites and their potential impact on organism health.

Data availability. The sequence of *Microcystis aeruginosa* PMC 728.11 has been deposited in GenBank under the BioProject accession number PRJNA650216 (GenBank accession number JADCRC010000000, BioSample number SAMN15699246, and SRA numbers SRR12746604 and SRR12746605). Strain PMC 728.11 is available from the collection of *Cyanobacteria* and Microalgae (PMC-ALCP) located in the Muséum National d'Histoire Naturelle (Paris, France; https://mcam.mnhn.fr/fr/collection-de-cyanobacteries-et-microalgues-vivantes-pmc-alcp-470).

REFERENCES

1. Harke MJ, Steffen MM, Gobler CJ, Otten TG, Wilhelm SW, Wood SA, Paerl HW. 2016. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp. Harmful Algae 54:4–20. https://doi.org/10.1016/j.hal.2015.12.007.

2. Humbert J-F, Barbe V, Latifi A, Gugger M, Calteau A, Coursin T, Lajus A, Castelli V, Oztas S, Samson G, Longin C, Medigue C, de Marsac NT. 2013. A tribute to disorder in the genome of the bloom-forming freshwater Cyanobacterium *Microcystis aeruginosa*. PLoS One 8:e70747. https://doi.org/10.1371/journal.pone.0070747.

3. Le Manach S, Sotton B, Huet H, Duval C, Paris A, Marie A, Yéprémian C, Catherine A, Mathéron L, Vinh J, Edery M, Marie B. 2018. Physiological effects caused by microcystin-producing and non-microcystin producing *Microcystis aeruginosa* on medaka fish: a proteomic and metabolomic study on liver. Environ Pollut 234:523–537. https://doi.org/10.1016/j.envpol.2017.11.011.

4. Duperron S, Halary S, Habiballah M, Gallet A, Huet H, Duval C, Bernard C, Marie B. 2019. Response of fish gut microbiota to toxin-containing cyanobacterial extracts: a microcosm study on the medaka (*Oryzias latipes*). Environ Sci Technol Lett 6:341–347. https://doi.org/10.1021/acs.estlett.9b00297.

5. Le Manach S, Duval C, Marie A, Djediat C, Catherine A, Edery M, Bernard C, Marie B. 2019. Global metabolomic characterizations of *Microcystis* spp. highlights clonal diversity in natural bloom-forming populations and expands metabolite structural diversity. Front Microbiol 10:791. https://doi.org/10.3389/fmicb.2019.00791.

6. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. Microbiology 111:1–61. https://doi.org/10.1099/00221287-111-1-1.

7. Andrews S. 2014. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

8. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.

9. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.

10. Bankevich A, Nurk S, Antipov D, Gurevich A. a, Dvorkin M, Kulikov AS, Lesin V, Nikolenko S, Pham S, Prjibelski A, Pyshkin A, Sirotkin A, Vyahhi N, Tesler G, Alekseyev M, Pevzner P. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.

11. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

12. Cambuy D, Coutinho F, Dutilh B. 2016. Contig annotation tool CAT robustly classifies assembled metagenomic contigs and long sequences. bioRxiv https://doi.org/10.1101/072868.

13. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829. https://doi.org/10.1093/bioinformatics/bts252.

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. Vallenet D, Engelen S, Mornico D, Cruveiller S, Fleury L, Lajus A, Rouy Z, Roche D, Salvignol G, Scarpelli C, Médigue C. 2009. MicroScope: a platform for microbial genome annotation and comparative genomics. Database (Oxford) 2009:bap021.

16. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NssH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132. https://doi.org/10.1186/s13059-016-0997-x.

17. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema M, Weber T. 2019. AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.

18. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG, Weber T, Medema MH. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458. https://doi.org/10.1093/nar/gkz882.

19. Simonsen M, Mailund T, Pedersen CN. (2008). Rapid neighbour-joining, p 113–122. *In* International workshop on algorithms in bioinformatic. Springer, Berlin, Germany. https://doi.org/10.1007/978-3-540-87361-7_10.

20. Demay J, Bernard C, Reinhardt A, Marie B. 2019. Natural products from cyanobacteria: focus on beneficial activities. Mar Drugs 17:320. https://doi.org/10.3390/md17060320.