



Complete Genome Sequence of Subcluster 5.2 *Synechococcus* sp. Strain CB0101, Isolated from the Chesapeake Bay

Daniel Fucich,^a David Marsan,^a Ana Sosa,^a Feng Chen^a

^aThe Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, Maryland, USA

ABSTRACT *Synechococcus* sp. strain CB0101 is a model strain for cyanobacteria living in the estuarine environment. It is also a representative member of marine *Synechococcus* subcluster 5.2. The draft genome sequence of CB0101 was reported in 2014 with 454 sequencing. Here, we report the complete genome sequence of CB0101, obtained with PacBio sequencing. CB0101 contains a specialized array of genes which are involved in sensing, responding to, and persisting in the presence of environmental stress.

Synechococcus is the ubiquitous genus of the picocyanobacteria. *Synechococcus* spp. contribute significantly to global primary productivity due to their high abundance (1). Many genomes of marine *Synechococcus* spp. have been reported (2–4), but the majority of them are coastal and oceanic *Synechococcus* spp., and few are from estuaries. It has been reported that a unique genotype and ecotype of *Synechococcus* are present in the Chesapeake Bay estuary (5–7). *Synechococcus* sp. strain CB0101 was isolated from the Chesapeake Bay with preincubation and pour plating using SN medium with adjusted salinity to match water samples described by Chen et al. (5). *Synechococcus* sp. CB0101 was grown in the SN medium at 25°C in constant light (20 to 30 $\mu\text{E m}^{-2} \text{s}^{-1}$) in an illuminated incubator. *Synechococcus* CB0101 is a representative strain of estuarine *Synechococcus* spp. which belongs to subcluster 5.2 (8, 9). CB0101 has shown much higher tolerance to temperature, salinity, and heavy metals than have coastal and oceanic *Synechococcus* strains (10). It appears that estuarine *Synechococcus* strains exhibit unique ecophysiological features that enable them to adapt to highly dynamic environments like the Chesapeake Bay. Because of its ecological relevance, the draft genome sequence of CB0101 was reported by Marsan et al. in 2014 (8). The draft genome sequence of CB0101 is 2,686,395 bp long, with 3,109 genes and 46 tRNAs and 15 total rRNA gene copies. Annotation yielded 3 5S rRNAs, 7 23S rRNAs, and 5 16S rRNAs. In order to fully understand the molecular mechanisms in response to environmental changes, we began to apply different omics (i.e., transcriptomics and proteomics) to the study of CB0101. We quickly noticed the need to obtain a complete genome sequence of CB0101.

The complete genome sequence of CB0101 was determined using a hybrid approach combining 454 GS-FLX Titanium 8-kb paired-end and PacBio libraries. Cells were collected using centrifugation (10,000 $\times g$, 10 min), and the pellet was transferred into a 2-ml tube and processed immediately. We obtained DNA from 25-ml cultures using the Epicentre MasterPure kit. The first round of sequencing was completed using a 454 GS FLX machine for whole-genome sequencing with the source being genomic. Selection was random with a paired-read layout. This yielded 439,486 reads with an average length of 445 bp. Unfortunately, only a draft genome could be completed using this approach, so to amend this and close the genome, PacBio sequencing at the Institute of Genome Sciences (IGS) obtained 121,716 total reads with an average read length of 6,295 bp, resulting in $\sim 40\times$ coverage on the genome. Parameters used for

Citation Fucich D, Marsan D, Sosa A, Chen F. 2019. Complete genome sequence of subcluster 5.2 *Synechococcus* sp. strain CB0101, isolated from the Chesapeake Bay. Microbiol Resour Announc 8:e00484-19. <https://doi.org/10.1128/MRA.00484-19>.

Editor Irene L. G. Newton, Indiana University, Bloomington

Copyright © 2019 Fucich et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Feng Chen, chenf@umces.edu.

D.F. and D.M. contributed equally to the article.

Received 2 May 2019

Accepted 6 August 2019

Published 29 August 2019

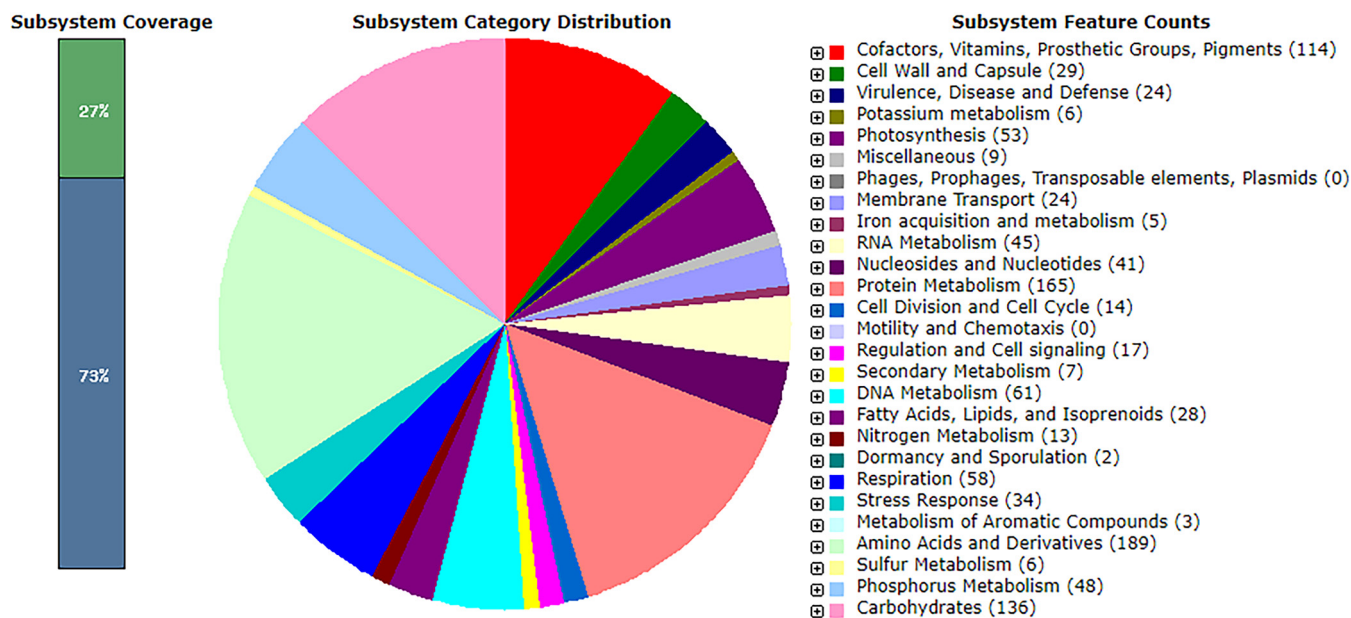


FIG 1 Subsystem coverage and distribution of *Synechococcus* sp. strain CB0101 (based on RAST annotation).

PacBio sequencing included a long library type (5 to 20 kb, polymerase P6, C4 sequencing chemistry, and a movie length of 240 min). The PacBio sequencing greatly improved genome assembly compared to the previous 454 sequencing run (8). Low-quality regions of sequencing data were removed from the raw PacBio reads using the Hierarchical Genome Assembly Process (HGAP) v2.2 assembly program, with default parameters (11). Contigs were ordered into putative scaffolds based on their similarity to closely related closed *Synechococcus* genomes, as determined using Mauve (12). As the cultures sequenced were known to contain heterotrophs, we identified the most “*Synechococcus*-like” contigs from nonaxenic cultures by searching each resulting contig against a custom database of sequenced marine microbial genomes using BLAST (13). Contigs with a best match to a non-*Synechococcus* genome were removed from the assembly. The remaining contigs, from both the 454 and PacBio sequencing runs, were combined and assembled using the Sequencher v5.3 software (Gene Codes Corporation, Ann Arbor, MI, USA). The combination of PacBio and 454 sequencing enabled us to assemble a complete genome of CB0101.

The complete genome of CB0101 consists of 2,789,657 bp (64.1% G+C content), with 3,128 genes, 52 tRNAs, and 24 rRNAs. The final assembly of CB0101 was annotated using the Rapid Annotations using Subsystems Technology (RAST) server (14–16). CB0101 contains many unique genes, as nearly half (46%) of the coding sequences (CDS) were hypothetical proteins. About 73% of the CDS (2,304 genes) were not grouped in the RAST functional subsystems (Fig. 1). CB0101 contains genes associated with an increased capacity to sense and respond to changes in the environment, including 24 membrane transporter genes, 17 regulation and cell signaling genes, 34 stress response genes, 24 virulence, disease, and defense genes, and 2 dormancy genes. Seven pairs of toxin-antitoxin (TA) genes were identified in the CB0101 genome (9, 10). TA genes are not commonly seen in *Synechococcus* spp. living in coastal and oceanic water (10). It has been shown that some TA genes are actively expressed under oxidative stress caused by zinc toxicity or high light exposure (9). Information gleaned from the complete genome showed the ability of CB0101 to sense and respond to environmental conditions, such as nitrogen or phosphate depletion and zinc metal excess. *Synechococcus* CB0101 contains specialized genes for persisting in the presence of highly dynamic environmental stressors and can be used to further study the molecular ecology of estuarine *Synechococcus* spp.

Data availability. This complete genome sequencing project has been submitted to the NCBI under the accession number [CP039373](https://doi.org/10.1093/mra/CP039373). The raw sequence reads for both the 454 GS-FLX Titanium 8-kb paired-end and PacBio libraries are available at accession numbers [SRX018027](https://doi.org/10.1093/mra/SRX018027) and [PRJNA529695](https://doi.org/10.1093/mra/PRJNA529695), respectively.

ACKNOWLEDGMENTS

The isolation, characterization, and genomic DNA preparation of CB0101 were supported by National Science Foundation grant MCB-0132070. The *Synechococcus* sp. strain CB0101 genome project was initiated with the support of the Gordon and Betty Moore Foundation.

The draft genome sequencing was performed at the J. Craig Venter Institute. The PacBio sequencing was conducted at IGS with additional funding from the Institute of Marine and Environmental Technology, UMCES.

REFERENCES

- Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, Karl D, Li W, Lomas M, Veneziano D, Vera C, Vrugt JA, Martiny AC. 2013. Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci U S A* 110:9824–9829. <https://doi.org/10.1073/pnas.1307701110>.
- Palenik B, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P, Lamerdin J, Regala W, Allen EE, McCarren J, Paulsen I, Dufresne A, Partensky F, Webb EA, Waterbury J. 2003. The genome of a motile marine *Synechococcus*. *Nature* 424:1037–1042. <https://doi.org/10.1038/nature01943>.
- Palenik B, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH, Madupu R, Nelson WC, Brinkac LM, Dodson RJ, Durkin AS, Daugherty SC, Sullivan SA, Khouri H, Mohamoud Y, Halpin R, Paulsen IT. 2006. Genome sequence of *Synechococcus* CC9311: insights into adaptation to a coastal environment. *Proc Natl Acad Sci U S A* 103:13555–13559. <https://doi.org/10.1073/pnas.0602963103>.
- Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, Tandeau de Marsac N, Wincker P, Dossat C, Ferreira S, Johnson J, Post AF, Hess WR, Partensky F. 2008. Unravelling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 9:R90. <https://doi.org/10.1186/gb-2008-9-5-r90>.
- Chen F, Wang K, Kan J, Bachoon DS, Lu J, Lau S, Campbell L. 2004. Phylogenetic diversity of *Synechococcus* in the Chesapeake Bay revealed by ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (rbcl) sequences. *Aquat Microb Ecol* 36:153–164. <https://doi.org/10.3354/ame036153>.
- Chen F, Wang K, Kan J, Suzuki MT, Wommack KE. 2006. Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* 72:2239–2243. <https://doi.org/10.1128/AEM.72.3.2239-2243.2006>.
- Wang K, Wommack KE, Chen F. 2011. Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. *Appl Environ Microbiol* 77:7459–7468. <https://doi.org/10.1128/AEM.00267-11>.
- Marsan D, Wommack KE, Ravel J, Chen F. 2014. Draft genome sequence of *Synechococcus* sp. strain CB0101, isolated from the Chesapeake Bay estuary. *Genome Announc* 2:e01111-13. <https://doi.org/10.1128/genomeA.01111-13>.
- Marsan D, Place A, Fucich D, Chen F. 2017. Toxin-antitoxin systems in estuarine *Synechococcus* strain CB0101 and their transcriptomic responses to environmental stressors. *Front Microbiol* 8:1213. <https://doi.org/10.3389/fmicb.2017.01213>.
- Marsan DW. 2016. Adaptive mechanisms of an estuarine *Synechococcus* based on genomics, transcriptomics, and proteomics. PhD thesis. University of Maryland, College Park, MD.
- Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
- Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>.
- McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Cowley AP, Lopez R. 2013. Analysis tool Web services from the EMBL-EBI. *Nucleic Acids Res* 41:W597–W600. <https://doi.org/10.1093/nar/gkt376>.
- 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. <https://doi.org/10.1186/1471-2164-9-75>.
- Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
- Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: a modular and extensible implementation of the RAST algorithm for annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.