Genome Sequence of the Estuarine *Synechococcus* sp. Strain NB0720_010

Campbell A. MacKenzie,[a] Marcia F. Marston,[b] Javier F. Tabima,[a] Nathan A. Ahlgren[a]

[a]Biology Department, Clark University, Worcester, Massachusetts, USA
[b]Department of Biology, Marine Biology and Environmental Science, Roger Williams University, Bristol, Rhode Island, USA

ABSTRACT Marine *Synechococcus* spp. are unicellular cyanobacteria widely distributed in the world's oceans. We report the complete genome sequence of *Synechococcus* sp. strain NB0720_010, isolated from Narragansett Bay, Rhode Island. NB0702_10 has several large (>3,000-amino acid) protein-coding genes that may be important in its interactions with other cells, including grazers in estuarine habitats.

Marine *Synechococcus* spp. are photosynthetic bacteria found in the surface waters of nearly every marine habitat on the planet and are estimated to contribute to ~17% of annual net marine primary productivity (1). While there are many genomes available for open-ocean *Synechococcus* spp. (2, 3), there has been limited genome sequencing of coastal and estuarine isolates (4). The genome sequence of *Synechococcus* sp. strain NB0720_010, isolated from Narragansett Bay and belonging to the marine *Synechococcus* subcluster 5.2 clade CB5, adds to our knowledge of genomic diversity in nearshore and estuarine *Synechococcus* populations.

Strain NB0720_010 was isolated from water collected on 28 July 2020 at 10:15 a.m. on an incoming tide in Narragansett Bay from a dock at 41°38′59.5812″N, 71°15′24.2038″W. NB0720_010 was isolated by pour plating using the methods outlined in reference 5 with modifications. Sample seawater was prefiltered through a 1-μm filter, mixed with an equal volume of 2× PRO99 medium (5) with 0.64% molten (~33℃) low-melting-point agarose, and immediately poured onto a 0.7% agarose PRO99 base layer. The plate was incubated at 22℃ with a 14:10-h light/dark illumination cycle at ~20 μE m^{-2} s^{-1} for several weeks until a single colony was picked and transferred to Pro99 liquid medium. After a few transfers to liquid medium, DNA was extracted from a late-exponential-phase culture using a phenol-chloroform extraction protocol (6, 7).

Illumina shotgun sequencing performed by the Microbial Genome Sequencing Center (Pittsburgh, PA; library preparation using the Nextera XT kit following the manufacturer's directions; sequencer, NextSeq 550) yielded 2,649,963 paired-end 151-bp reads. Default parameters were used for all software unless otherwise specified. The reads were trimmed using Trimmomatic v0.38 (8) with the following settings: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10, LEADING:10, TRAILING:10 SLIDINGWINDOW:4:15 MINLEN:50. Long-read Nanopore sequencing was performed in-house using DNA from the same culture but extracted from a different inoculum (library preparation using the SQKLSK-110 ligation kit following the manufacturer's directions; Flongle flow cell R9.4.1), and this yielded 874,048 reads (mean length, 255 bp; N_{50}, 69,887 bp; base calling using Guppy v5.0.11 + 2b6dbff; Oxford Nanopore). Assembly of the Illumina and Nanopore reads using Unicycler v0.4.9b (9) generated 630 contigs (N_{50}, 8 kb; total length, 18 Mb). Among these was a 2,410,448-bp circular contig with 63% GC content and 97.1% average nucleotide identity (calculated using FastANI [10]) to estuarine *Synechococcus* sp. strain CB0205 (GenBank accession number GCA_000179255.1) (11, 12), belonging to subcluster 5.2 clade CB5, and thus, this contig was deemed to be the complete genome of *Synechococcus* sp. NB0720_010. It was annotated using the NCBI

TABLE 1 Characteristics of the four very large (>3,000-aa) proteins in *Synechococcus* sp. strain NB0720_010

Locus tag	Protein length (aa)	Similarity to other giant proteins in *Synechococcus* isolate genomes[a]	Domain description
LY254_04120	10,749	30% identity over 44% of the protein to SynWH8101_0818 (4,083 aa; QBE68408.1) in *Synechococcus* sp. strain WH 8101 (16)	Several repeated domains (~100 aa long) with similarity to cadherin-like domains (pfam17803, pfam17892), bacterial Ig[b] domains (pfam17963), and repeats (TIGR01965) found frequently in *Vibrio*, *Colwellia*, *Bradyrhizobium*, and *Shewanella* (VCBS) members
LY254_03050	4,609	30% identity over 82% of the protein to EVJ50_06065 (11,376 aa; QEY31874.1) in *Synechococcus* sp. strain RSCCF101 (17)	Several repeated domains (DUF5801, pfam19116)
		36% identity over 82% of the protein to *Legionella pneumophila* toxin protein RtxA (STX70670.1)	
LY254_05360	3,090	No hits to isolate genomes but several to short (<700-aa) proteins from *Synechococcaceae* freshwater metagenomic assembled genomes (18)	Several repeated cadherin-like domains (pfam17892) and a C-type lectin-like domain (cd03603)
LY254_11890	3,087	96% identity over 100% of the protein to KJJ24_00540 (3,087 aa; QVV67743.1) in *Synechococcus* sp. strain LA31 (Narragansett Bay isolate) (4)	RTX toxin domain repeats (NF033203, NF033943) and an RTX toxin-related domain (COG2931)

[a] Top hit(s) from a *Synechococcus* isolate from a blastp search (13) against NCBI's nonredundant (nr) database with an E value of <1e−5. The lengths and GenBank accession numbers of the protein hits are listed in parentheses.

[b] Ig, immunoglobin.

Prokaryotic Genome Annotation Pipeline v5.3 (13), yielding 2,530 protein-coding genes, 3 complete rRNA operons (5S, 16S, and 23S), and 46 tRNAs.

NB0720_010 contains several putative type II antitoxin/toxin protein-coding genes. Most of these have homologs (determined using a blastp search [14]; E value < 1e−10) in other estuarine but not open-ocean *Synechococcus* isolate genomes, consistent with previous observations (15). NB0720_010 also contains four very large (>3,000 amino-acid [aa]), or "giant," protein-coding genes that all have repeated motifs possessing similarity to cadherin-like or RTX toxin domains (Table 1). Such giant proteins often occur in marine *Synechococcus* genomes (19), including SwmB in *Synechococcus* sp. WH8102 (10,791 aa), which is involved in swimming motility and resisting protistan predation (19–21). Similar to SwmB, we suggest that these giant proteins in NB0720_010 may be important for defense and/or competition, which may be crucial in estuarine waters, which typically have higher cell densities than open-ocean regions.

Data availability. *Synechococcus* sp. NB0720_10 is available from Nathan A. Ahlgren upon request. The sequence data are available at NCBI under BioProject accession number PRJNA793027, including the raw reads (SRA accession numbers SRR18042579 and SRR18042580) and the assembled genome (GenBank accession number CP090898).

REFERENCES

1. Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, Karl DM, Li WKW, Lomas MW, Veneziano D, Vera CS, 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.

2. Lee MD, Ahlgren NA, Kling JD, Walworth NG, Rocap G, Saito MA, Hutchins DA, Webb EA. 2019. Marine *Synechococcus* isolates representing globally abundant genomic lineages demonstrate a unique evolutionary path of genome reduction without a decrease in GC content. Environ Microbiol 21:1677–1686. https://doi.org/10.1111/1462-2920.14552.

3. Garczarek L, Guyet U, Dore H, Farrant GK, Hoebeke M, Brillet-Gueguen L, Bisch A, Ferrieux M, Siltanen J, Corre E, Le Corguille G, Ratin M, Pitt FD, Ostrowski M, Conan M, Siegel A, Labadie K, Aury JM, Wincker P, Scanlan DJ, Partensky F. 2021. Cyanorak v2.1: a scalable information system dedicated to the visualization and expert curation of marine and brackish picocyanobacteria genomes. Nucleic Acids Res 49:D667–D676. https://doi.org/10.1093/nar/gkaa958.

4. Kling JD, Webb EA, Hutchins DA. 2022. Genome sequence of *Synechococcus* sp. strain LA31, isolated from a temperate estuary. Microbiol Resour Announc 11:e0077521. https://doi.org/10.1128/mra.00775-21.

5. Moore LR, Coe A, Zinser ER, Saito MA, Sullivan MB, Lindell D, Frois-Moniz K, Waterbury J, Chisholm SW. 2007. Culturing the marine cyanobacterium *Prochlorococcus*. Limnol Oceanogr Methods 5:353–362. https://doi.org/10.4319/lom.2007.5.353.

6. Lever MA, Torti A, Eickenbusch P, Michaud AB, Santl-Temkiv T, Jorgensen BB. 2015. A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. Front Microbiol 6:476. https://doi.org/10.3389/fmicb.2015.00476.

7. Thornton CN, Brazelton WJ. 2020. DNA extraction from Sterivex filters. Protocolsio. https://doi.org/10.17504/protocols.io.bm2ek8be.

8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

9. 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.

10. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.

11. Fucich D, Xu Y, Sosa A, Jia Y, Zhang R, Jiao N, Chen F. 2021. Complete genome sequences of Chesapeake Bay *Synechococcus* strains CBW1002 and CBW1006 isolated in winter. Genome Biol Evol 13:evab009. https://doi.org/10.1093/gbe/evab009.

12. Chen F, Wang K, Kan J, Suzuki M, Wommack K. 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.

13. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

14. Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. Biol Direct 7:12. https://doi.org/10.1186/1745-6150-7-12.

15. Fucich D, Chen F. 2020. Presence of toxin-antitoxin systems in picocyanobacteria and their ecological implications. ISME J 14:2843–2850. https://doi.org/10.1038/s41396-020-00746-4.

16. Marston MF, Polson SW. 2020. Whole-genome sequence of the cyanobacterium *Synechococcus* sp. strain WH 8101. Microbiol Resour Announc 9:e01593-19. https://doi.org/10.1128/MRA.01593-19.

17. Ng YM. 2019. Characterization of Red Sea cyanobacteria aimed for cell factory applications in Saudi Arabia: *Synechococcus* sp. RSCCF101. PhD dissertation. King Abdullah University of Science and Technology, Thuwal, Saudi Arabia.

18. Rodriguez-R LM, Tsementzi D, Luo C, Konstantinidis KT. 2020. Iterative subtractive binning of freshwater chronoseries metagenomes identifies over 400 novel species and their ecologic preferences. Environ Microbiol 22:3394–3412. https://doi.org/10.1111/1462-2920.15112.

19. Dufresne A, Ostrowski M, Scanlan D, Garczarek L, Mazard S, Palenik B, Paulsen I, de Marsac N, Wincker P, Dossat C, Ferriera S, Johnson J, Post A, Hess W, Partensky F. 2008. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. Genome Biol 9:R90. https://doi.org/10.1186/gb-2008-9-5-r90.

20. Strom SL, Brahamsha B, Fredrickson KA, Apple JK, Gutierrez Rodriguez A. 2012. A giant cell surface protein in *Synechococcus* WH8102 inhibits feeding by a model dinoflagellate predator. Environ Microbiol 14:807–816. https://doi.org/10.1111/j.1462-2920.2011.02640.x.

21. McCarren J, Brahamsha B. 2007. SwmB, a 1.12-megadalton protein that is required for nonflagellar swimming motility in *Synechococcus*. J Bacteriol 189:1158–1162. https://doi.org/10.1128/JB.01500-06.