The large genome of Synechococcus moorigangaii CMS01 isolated from a mangrove ecosystem - evidences of motility and adaptive features

Anwesha Ghosh1 and Punyasloke Bhadury1,2

1 Centre for Climate and Environmental Studies, Indian Institute of Science Education and Research Kolkata, Mohanpur-741246, Nadia, West Bengal, India
2 Integrative Taxonomy and Microbial Ecology Research Group, Department of Biological Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur-741246, Nadia, West Bengal, India

E-mail: pbhadury@iiserkol.ac.in

Keywords: Synechococcus, Sundarbans, coastal ocean, Bay of Bengal, estuary

Abstract

The whole genome of Synechococcus moorigangaii CMS01 isolated from Indian Sundarbans mangroves of Bay of Bengal is about 5.5 Mbp in size and contains approximately 0.5 Mbp plasmids. Genome annotation revealed total of 5806 genes out of which 5701 were CDSs. Of these, 5616 coding genes with 5616 protein coding CDSs were found. Along with genes coding for essential metabolic proteins, transport proteins and other cellular apparatus, genome also codes for proteins involved in flagella and pilus formation which has not been widely reported before in any coastal species of Synechococcus. The genome contains one incomplete prophage sequence. The genome analysis revealed adaptive features of S. moorigangaii CMS01 and establishes its ubiquitous distribution in coastal water of Bay of Bengal.

Introduction

The picocyanobacterial taxon Synechococcus represented by numerous species is unicellular, coccoid or rod-shaped with ubiquitous distribution in marine ecosystems [1]. Synechococcus (>3 μm in diameter) have the ability to nutrients at sub-micromolar concentrations [2]. Members of this genus can grow across a range of light intensities and spectral range [3] and can utilize variety of nitrogen sources for growth [4]. Synechococcus belongs to a polyphyletic group which is characterized by genes acquired through horizontal gene transfer (HGT). The process of HGT plays a crucial role in evolution of cyanobacterial genomes [5] and can be identified as genomic islands within the genome. Genomic islands are thought to be acquired during infection by cyanophages [6]. Such genes acquired by HGT can code for proteins involved in photosynthesis, metabolism of key elements such as carbon and phosphorus and other stress responses [7].

We describe the genome sequence of a previously described new species of Synechococcus, named Synechococcus moorigangaii CMS01 [8] and highlight some adaptive features that has resulted in its ubiquity in coastal water of Bay of Bengal.

Methods

Isolation and genomic DNA extraction

Synechococcus moorigangaii CMS01 was isolated and previously identified [8] from estuarine surface water of Stn3 of Sundarbans Biological Observatory Time Series (SBOTS) located in the Mooriganga estuary of Indian Sundarbans facing the coastal Bay of Bengal. Cells were isolated on ASN III (+N) medium of salinity 15. The
culture was grown for 15 days under continuous light. Genomic DNA (gDNA) was extracted using modified published protocol [9].

**Whole-genome sequencing**

Genome sequencing library was generated using Illumina-compatible SureSelect\textsuperscript{XT} whole genome preparation kit (Agilent, USA), followed by amplification and sequencing on Illumina MiSeq platform. The sequence data was checked using FastQC and adapters were trimmed using Cutadapt [10]. Quality checked pair-end reads were assembled into contigs using Unicycler [11].

**Whole-genome sequence annotation and comparisons**

The genome of *S. moorigangaii* CMS01 was aligned into circular map using CGView server [12]. Genome sequence of *S. moorigangaii* CMS01 along with *Synechococcus* sp. PCC 7003 and *Synechococcus* sp. PCC 7335 were aligned in progressiveMauve [13]. The whole genome sequence based phylogeny was performed in the Type (Strain) Genome Server (TYGS) [14]. The genome data was compared using the MASH algorithm. Genome distances were calculated using the Genome BLAST Distance Phylogeny (GBDP) approach. The resulting inter-genomic distances were used to calculate a balance minimum evolution tree with branch support via FASTME 2.1.4 including SPR post-processing [15]. Branch support was inferred from 100 pseudo-bootstrap replicates.

Genomic relatedness with closest relatives was determined using OrthoANIu algorithm [16]. Digital DDH values were calculated using genome-genome distance calculator (GGDC 2.1) applying Formula 2 (identities/ HSP length) [17]. The genome sequences used for GGDC and OrthoANIu analyses were of *Synechococcus* sp. PCC 7335, *Synechococcus* sp. PCC 7003, *Synechococcus* sp. PCC 7002 and *Synechococcus* sp. NKBG15041c.

Average amino acid index (AAI) was determined using AAI-profiler [18]. *In silico* phenotyping was performed using Traitar [19]. Genomic annotation was carried out using Prokka [20] and revalidated using the Prokaryotic Genome Annotation Pipeline (PGAP) [21]. Genomic islands were predicted using IslandViewer 4 [22]. The resulting protein profile was viewed by plotting the data in a circular map using GView [23]. Prophage sequences within the genome and plasmids were identified using PHASTER [24, 25]. The accession number for submitted genome data is SAMN12191289.

**Results and discussion**

**Genome analysis**

The draft genome of *S. moorigangaii* CMS01 consisted of 6066887 bases which assembled into 227 contigs (figure 1). Approximately, 500000 bases were identified as plasmids. The genome is about 5.5 Mbp in size and nearly twice as big as the closest relatives including *Synechococcus* sp. PCC 7002, *Synechococcus* sp. PCC 7003 and *Synechococcus* sp. NKBG15041c (figure S1 available online at stacks.iop.org/IOSPN/1/034001/mmedia). The GC content was 56.53%. Genome analysis indicated the presence of 5806 total genes out of which 5701 were CDSs. A total of 5616 coding genes with 5616 protein coding CDSs were found. The genome codes for 105 RNA pseudogenes and 1 CRISPR array were found. Based on whole genome phylogeny *S. moorigangaii* CMS01 was confirmed as a new species (figure S3). GGDC(%) and orthoANIu(%) also confirmed *S. moorigangaii* CMS01 as a new species (table 1). AAI analysis showed maximum amino acid identity with *Synechococcus* sp. PCC 7002 (Average AAI = 0.875) (table 1).

**Possible phenotypic traits from genotype**

**Metabolism**

*In silico* phenotyping indicated the organism to be aerobic, motile and Gram negative. The isolate is susceptible to bile and produces enzymes including casein hydrolyase, arginine dihydrolase, alkaline phosphatase, oxidase, catalase, lipase, lysine decarboxylase, gelatinase, coagulase, urease and DNase. It can use glycerol, pyrorolidonyl-beta-naphthylamide, D-mannitol, acetate, L-arabinose, mucate and tartrate. Growth utilizing carbon sources including sucrose, D-mannose and trehalose was found. It can convert nitrite to nitrogen gas. It can grow on MacConkey agar in presence of high NaCl concentration (figure S2).

**Energy generation and nutrient uptake**

Some of the genes identified from both the positive and negative strands of the gDNA are shown in figure S4. Genome annotation revealed genes involved in photosynthesis including photosystem II reaction centre proteins (psbN, psbH, psbL, psbI, psbZ, psbQ), photosystem I iron-sulfur centre protein (psaC), photosystem I reaction centre subunits VIII, IX, and nitrogen regulation including Mo-dependent nitrogenase and global nitrogen regulator (*ntrA*). Genes involved in arsenic regulation including arsenic-transporting ATPase, ACR3...
family arsenite efflux transporter, arsenic resistance protein (arsH) and arsenate reductase (arsC) were found. Other genes involved in toxin systems including type II toxin-antitoxin system (vapC) were identified. Genes involved in iron regulation including thioredoxin (trxA) were found. Transport protein coding genes including TonB-dependent receptor and MFS transporter were identified. Transporters specific for uptake of urea including urtABCDE transport system and urease cluster ureABCDEFG were present. A large number of cyanobacterial members harbour genes encoding urea catabolytic urease (ureABCDEFG), but not along with urea transport system [26]. The isolate can grow in presence of urea (1 μM concentration) although nitrate is the preferred source for growth [8]. A large number of genes coding for phosphate/phosphite/phosphonate ABC transporter indicate the capacity to uptake phosphorus in different forms from the environment. Indeed in estuarine mangroves phosphate can be limiting [27] and these genes reflect metabolic ability to uptake available forms of phosphorus.

Environmental adaptations
Transporter-coding genes indicated the possible adaptive capability of S. moorigangaii CMS01 to cope with extreme conditions including temperature and salinity which vary seasonally in Sundarbans leading to its ubiquitous distribution [28]. Molecules such as glutathione help to maintain cell redox homeostasis and protect the cell membrane lipids from oxidation stress in cold conditions [29]. The genome codes for RpsB, RpsC, RpsR, RpsG, RpsH, RpsI, RpsJ, RpsK, RpsM, RpsN, RpsO, RpsP, RpsQ and RpsS proteins. The genome codes for linker polypeptides that are necessary for correct assembly of phycobiliprotein in phycobilisome rods [2].

Table 1. Comparison of Synechococcus moorigangaii CMS01 with closest relatives.

| Organism                        | GGDC % | OrthoANIu % | Average AAI |
|---------------------------------|--------|-------------|-------------|
| Synechococcus sp. PCC 7002      | 0.1905 | 80.60       | 0.875       |
| Synechococcus sp. PCC 7003      | 0.1905 | 80.60       | 0.874       |
| Synechococcus sp. NKBG15041c    | 0.1975 | 79.61       | —           |
| Synechococcus elongatus PCC 7942| 0.1620 | 67.50       | —           |

Figure 1. Genome map of Synechococcus moorigangaii CMS01 in comparison with the closest relatives. The circular map also shows GC content and GC skew (+/−) of S. moorigangaii CMS01. The gap portions show no overlapping regions with closest neighbours.
Phycobilisome rod-core linker polypeptide (CpcG) was detected. Phycobilisome degradation protein (NblA) and NblA-related protein were detected which function in the degradation of phycobilisomes during nutrient stress in cyanobacteria [30]. The genome codes for phycocyanin alpha and beta subunits along with allophycocyanin. Multiple copies of smpB gene were identified in genome that codes for proteins required rescuing ribosomes stalled on defective messages [31]. The genes associated with possible estuarine adaptations of this species have been summarized in table S1.

The genes possibly acquired by HGT as deduced by IslandViewer 4 are enlisted in table S2. The presence of multidrug efflux systems indicates the need to export toxins and highlight potential competition for resources experienced by S. moorigangaii CMS01. The possible locations of genomic islands are shown in figure 2. One incomplete prophage sequence of 14.6 kb length was identified. This sequence lays between positions 90307–105001 nucleotide position within the genome and codes for 18 proteins. This prophage sequence shows maximum identity with Paracoccus phage vB_PmaS-IMEP1 (Accession number: NC_026608).

**Motility**

Though not observed earlier by microscopy, genome data indicates motility in S. moorigangaii CMS01. Proteins including flagellar motor stator protein (MotA), flagellar motor switch protein (FliM, FliN), flagellar protein (FlaG), flagellar protein export ATPase (FliI), flagellar export protein (FliJ), flagellar basal-body protein (FlbY), flagellar basal body P-ring protein (FlgI) and formation protein (FlgA), flagellar basal body L-ring protein (FlgH), flagellar basal-body rod protein (FlgB, FlgC, FlgG, FlgF), flagellar basal body protein (FliL), flagellar hook-basebody complex (FlIE), flagellar biosynthesis protein (FlhA, FlhB, FlgI, FlIP), flagellar biosynthesis-like protein (FlhF), flagellar biosynthesis regulator (FlaF), flagellar type III secretion system protein (FltQ, FltR), flagellar type III secretion system pore protein (FltP) and flagellar biosynthesis repressor (FlbT) are encoded by the genome. In addition to flagella, the genome also codes for pilus formation proteins. These include Pilus assembly proteins (CpaA, CpaD, PilM), pilus assembly protein (CpaB) and type IV pilus twitching motility protein (PilT). These features have not been widely reported in coastal species of *Synechococcus*.

![Figure 2](image_url)
Conclusion

The genome of *S. moirangtangui* CMS01 is about 5.5 Mbp which is nearly double the size compared to closest neighbours. Sequence data indicated the presence of plasmids. Genome analysis reveals genes involved in uptake of various nitrogenous compounds including amino acids and urea and presence of a wide array of cold shock proteins. Genes coding for flagella and pilus formation were also detected. The genome contains genomic islands acquired by horizontal gene transfer which codes for specialized proteins including CRISPR associated proteins.

Acknowledgments

This work was supported by ARF grant of IISER Kolkata and partly supported by SwarnaJayanti Fellowship of Department of Science & Technology, Govt of India (DST/SJF/E&ASA-01/2017-18) awarded to Punyasloke Bhadury.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://www.ncbi.nlm.nih.gov

ORCID iDs

Anwesha Ghosh @ https://orcid.org/0000-0003-1765-9832

Punyasloke Bhadury @ https://orcid.org/0000-0001-8714-7475

References

[1] Flornhaun P et al 2013 Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus* Proc. Natl Acad. Sci. USA 110 9824–9

[2] Palenik B, Ibrahimshah B, Larimer F W, Land M, Hauser L, Chain P, Lamerdin J, Regala W, Allen E E and Mccarren J 2003 The genome of a motile marine *Synechococcus* Nature 424 1037–42

[3] Kana T M and Gilbert P M 1987 Effect of irradiances up to 2000 μE m-2 s-1 on marine *Synechococcus WH7803* - I. Growth, pigmentation, and cell composition Deep Sea Res. Part A 34 479–95

[4] Moore L R, Post A F, Rocap G and Chisholm S W 2002 Utilization of different nitrogen sources by the marine cyanobacteria *Prochlorococcus* and *Synechococcus* Limnof Oceano. 47 989–96

[5] Nakamura Y, Itoh T, Matsuha H, Gojobori T, Itoh T, Matsuha H, Gojobori T, Matsuha H, Gojobori T and Gojobori T 2004 Biased biological functions of horizontally transferred genes in prokaryotic genomes *Nat. Genet.* 36 760–6

[6] Palenik B, Ren Q, Tai V and Paulsen I T 2009 Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity *Environ. Microbiol.* 11 349–59

[7] Coutinho F, Tschoeke D A, Thompson F and Thompson C 2016 Comparative genomics of *Synechococcus* and proposal of the new genus *Parasynechococcus* PeerJ 4 e1522

[8] Singh T and Bhadury P 2019 Description of a new marine planktonic cyanobacterial species *Synechococcus moirangtangui* (Order *Chroococcales*) from Sundarbans mangrove ecosystem *Phytotaxa.* 392 265–77

[9] Boström K H, Simu K, Hagström A and Riemann L 2004 Optimization of DNA extraction for quantitative marine bacterioplankton community analysis *Limnol Oceanogr Methods.* 2 365–73

[10] Marcel M 2011 Cutadapt removes adapter sequences from high-throughput reads *EMBnet. Journal* 17 10–12

[11] Wick R R, Judd L M, Gorrie C L and Holt K E 2016 Unicycler: resolving bacterial genome assemblies from short and long sequencing reads *PLoS Comput. Bio.* 13 e1005395

[12] Grant J R and Stothard P 2008 The CGView server: a comparative genomics tool for circular genomes *Nucleic Acids Res.* 36 W181–4

[13] Darling A E, Mau B and Perna N T 2010 progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement *PLoS One* 5 e11147

[14] Meier-Kolthoff P J and Göker M 2019 TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy *Nat Comm.* 10 2182

[15] Lefort V, Desper R and Gascuel O 2015 FastMe 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program *Mol Biol. Evol.* 2015 32 2798–800

[16] Yoon S H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H and Chun J 2017 Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies *Int. J. Syst. Evol. Microbiol.* 67 1613–9

[17] Meier-Kolthoff P J, Auch A F, Klenk H-P and Goker M 2013 Genome sequence-based species delimitation with confidence intervals and improved distance functions *BMC Bioinf.* 14 160

[18] Medlar A J, Törönen P and Holm L 2018 AAI-profiler: fast proteome-wide exploratory analysis reveals taxonomic identity, misclassification and contamination *Nucleic Acid Res.* 46 W479–85

[19] Weimann A, Moore K, Frank J, Pope P B, Bremges A and McHardy A C 2016 From genomes to phenotypes: traitar, the Microbial Trait Analyzer mSysten. 1 e00101-16

[20] Seemann T 2014 Prokka: rapid prokaryotic genome annotation *Bioinformatics.* 30 2068–9

[21] Tatusova T et al 2016 NCBI prokaryotic genome annotation pipeline *Nucleic Acids Res.* 44 6614–24
[22] Bertelli C et al. 2017 IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 45 W30–5

[23] Petkau A, Stuart-Edwards M, Stothard P and Domselaar G V. 2010 Interactive microbial genome visualization with GView. Bioinformatics. 26 3125–6

[24] Zhou Y, Liang Y, Lynch K H, Dennis I J and Wishart D S. 2011 PHAST: a fast phage search tool. Nucleic Acids Res. 39 W347–52

[25] Arndt D, Grant J, Marcu A, Sajed T, Pon A, Liang Y and Wishart D S. 2016 PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44 W16–21

[26] Veaudour T, Cassier-Chauvat C and Chauvat F. 2019 Genomics of urea transport and catabolism in Cyanobacteria: biotechnological implications. Front Microbiol. 10 2052

[27] Choudhury A K, Das M, Philip P and Bhadury P. 2015 An assessment of the implications of seasonal precipitation and anthropogenic influences on a mangrove ecosystem using phytoplankton as proxies. Est Coast. 38 854–72

[28] P. Bhadury and Singh T. 2020 Analysis of marine planktonic cyanobacterial assemblages from Mooriganga Estuary, Indian Sundarbans using molecular approaches. Front Mar. Sci. 7 222

[29] Tanizawa Y, Tohno M, Kaminuma E, Nakamura Y and Arita M. 2015 Complete genome sequence and analysis of Lactobacillus hokkaidonensis LOOC260^T, a psychrotrophic lactic acid bacterium isolated from silage. BMC Genom. 16 240

[30] Dolganov N and Grossman A R. 1999 A polypeptide with similarity to phycocyanin α-subunit phycocyanobilin lyase involved in degradation of phycobilisomes. J. Bacteriol. 181 610–7

[31] Karzai A W, Susskind M M and Sauer R T. 1999 SmrB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). EMBO J. 18 3793–9