

Gene diversity and organization in *rbcL*-containing genome fragments from uncultivated *Synechococcus* in the Gulf of Mexico

David E. John¹, Boris Wawrik², F. Robert Tabita³, John H. Paul^{1,*}

¹College of Marine Science, University of South Florida, 140 7th Avenue S, St. Petersburg, Florida 33701, USA

²Biotechnology Center for Agriculture and the Environment, Foran Hall, Cook College, Rutgers — The State University of New Jersey, 59 Dudley Road, New Brunswick, New Jersey 08901, USA

³Department of Microbiology and Plant Molecular Biology/Biotechnology Program, The Ohio State University, Columbus, Ohio 43210, USA

ABSTRACT: Factors controlling phytoplankton distribution and activity in the marine environment are often poorly understood. Knowledge of genomic content and organization may aid in understanding these processes. To this end, bulk environmental DNA was extracted from Gulf of Mexico seawater and cloned into a bacterial artificial chromosome (BAC) vector to generate a library of 3000 large-insert clones. This library was probed for various types of RubisCO gene (*rbcL*); 8 *rbcL*-containing clones (20 to 73 kb in size) were detected and sequenced. *rbcL* sequences indicated that they originated from marine α -*Synechococcus* spp. Genomic organization was highly similar to that of the sequenced genome of *Synechococcus* sp. strain WH 8102 in the region surrounding *rbcL*. Several differences from known α -*Synechococcus* were also observed, including putative monoamine oxidase and amino acid transport genes, and electron transport, photosynthesis and hypothetical protein genes similar to those of *Prochlorococcus* spp. (i.e. flavodoxin, thioredoxin reductase and iron-stress-induced chlorophyll-binding proteins). These findings indicate that the order and composition of the carbon fixation operon (*cbbX* to *ccmK*) is highly conserved in α -*Synechococcus*, while adjacent genomic composition indicates expanded metabolic capabilities obtained from gene-transfer events with *Prochlorococcus* spp. and other cyanobacteria.

KEY WORDS: Metagenomics · *Synechococcus* · *rbcL* · Carbon fixation · Monoamine oxidase · Iron stress · *isiA* · Chlorophyll-binding proteins

—Resale or republication not permitted without written consent of the publisher—

INTRODUCTION

Most marine bacteria remain to be cultured, and a very small percentage of marine microorganisms are represented by isolates that have been cultivated and described (Schmidt et al. 1991, Sieracki & Viles 1992, Robertson et al. 1998, Button 2001). Although emerging techniques such as FISH (fluorescence *in situ* hybridization)-microautoradiography, isotope arrays, and stable isotope probing are allowing better coupling of metabolic capabilities with phylogenetic identity (Dumont & Murrell 2005), much of the metabolic diversity of uncultured microbes remains unknown.

Large-insert (metagenomic) clone libraries, using bacterial artificial chromosome (BAC), cosmid or fosmid vectors, have been employed for studying environmental microbial diversity and potential function from uncultivated natural communities without potential amplification biases. In several prior studies, desired clones were identified from libraries using ribosomal gene probes for phylogenetic anchors, and a metabolic function of the organism was inferred from interpretation of surrounding genes (Stein et al. 1996, Schleper et al. 1997, Beja et al. 2000b, Rondon et al. 2000, Sabehi et al. 2003). This approach led to the discovery of a novel phototrophy in the oceans, involving

*Corresponding author. Email: jpaul@marine.usf.edu

proteorhodopsin (Beja et al. 2000a). Subsequent metagenomic analysis indicated that rhodopsin-like proteins were abundant in the Sargasso Sea (Venter et al. 2004).

The *rbcL* gene, encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO), has been employed to determine the distribution and activity of marine phytoplankton (Wawrik et al. 2003, Corredor et al. 2004, Wawrik & Paul 2004), and may serve as a useful anchor for metagenomic investigations. RubisCO is the key enzyme for the reductive assimilation of CO₂ by the Calvin-Benson-Basham pathway, the primary mechanism for CO₂ fixation among plants, algae and autotrophic bacteria (Tabita 1988, 1999). RubisCO is believed to be responsible for as much as 95% of all carbon fixed in marine systems (Raven 1995).

The enzyme occurs in distinct forms, Forms I to IV (Tabita & Hanson 2004), in bacteria, archaea, and prokaryotic and eukaryotic phytoplankton, with the predominant Form I protein further subdivided into Forms IA, IB, IC, ID in different groups of phytoplankton and bacteria (Tabita 1988, 1995). In marine phytoplankton, Form II, usually associated only with certain bacteria, has been found to occur in dinoflagellate species, but otherwise Form IB or Form ID is the norm for these organisms. Form IA is found in nitrifying and sulfur-oxidizing chemoautotrophic bacteria as well as in most marine *Synechococcus* (the marine Type A or α -*Synechococcus*) and all *Prochlorococcus* strains sequenced. All other cyanobacteria (β -cyanobacteria) as well as all green algae possess the Form IB enzyme. Form IC RubisCO (encoded by the *cbbLS* genes) is used by some phototrophic and chemoautotrophic bacteria, while the Form ID enzyme is associated with a diverse group of eukaryotic lineages including essentially all chromophytic, eukaryotic algae, such as diatoms and prymnesiophytes. Form III RubisCO is found in many anaerobic archaea, which have been shown to synthesize catalytically active enzymes (Watson et al. 1999, Finn & Tabita 2003). In contrast, cryptic RubisCO-like protein (RLP) sequences (Form IV RubisCO) that do not catalyze ribulose bisphosphate carboxylation have been found in many bacteria (Hanson & Tabita 2001, 2003, Tabita & Hanson 2004) and at least 1 archaeal genome (Klenk et al. 1997). It should also be stressed that some authors consider *rbcL* gene sequences a more accurate marker for the classification of cyanobacteria than 16S rRNA gene sequences (Badger et al. 2002).

Analyzing the expression of *rbcL* genes may provide knowledge of the distribution and activity of phytoplankton in the marine environment; however, such information provides no indication as to what other genetic factors might relate to the relative abundance

or activity of such organisms. In order to study these factors, we employed metagenomic cloning to investigate the organization and diversity of genes in the region surrounding *rbcL*, which we term the '*rbcL* genomic environment,' in a phytoplankton population from the Gulf of Mexico.

MATERIALS AND METHODS

Sampling and DNA processing. Seawater sampling was performed aboard the RV 'F. G. Walton Smith' in the Gulf of Mexico; 240 l was collected by rosette sampler from a depth of 40 m at an oligotrophic station (25° 15.78' N, 84° 13.21' W). Cells were concentrated by vortex-flow filtration (Membrex) with a 100 kDa membrane filter, and then further concentrated by centrifugation at 18 500 × *g* for 15 min. Cell pellets were suspended in 1% low-melting-point agar (BioWhittaker Molecular Applications) and molded into 75 μ l plugs onboard ship. Plugs were extruded into lysis buffer (10 mM Tris × HCl pH 8.0; 50 mM NaCl, 0.1 M EDTA, pH 8.0; 1% sarcosyl; 0.2% sodium deoxycholate; 1 mg ml⁻¹ lysozyme; 1 ml per plug) and incubated at 37°C for 2 h. Lysis buffer was replaced with ESP buffer (1% sarcosyl; 1 mg ml⁻¹ Proteinase K; 0.5 M EDTA, pH 8.0) and incubated overnight at 55°C, after which time the buffer was replaced with fresh ESP without Proteinase K and incubated for an additional 60 min. Plugs were stored in 0.5 M EDTA, pH 8.0 at 4°C.

To digest DNA in agarose plugs, 5 plugs were transferred through the following sequential mixtures: (1) 20 ml 10% sarcosyl, 25 mM EDTA, 25 mM tris-HCl (both pH 8.0) for 2 h at 37°C; (2) 25 ml 20% ethanol with shaking for 1 h at 4°C; (3) 25 ml 25 mM tris-HCl (pH 8.0) for 1 h at 4°C; (4) 25 ml restriction digestion Buffer E (Promega) for 1 h at 4°C; (5) 5 ml 1× Buffer E, 4 mM spermidine, 50 μ l ml⁻¹ BSA, 160 U ml⁻¹ *Hind*III enzyme (Promega) overnight at 4°C to diffuse the enzyme into plugs; (6) warmed to 37°C, MgCl₂ added to 6 mM final concentration, incubated 8 h at 37°C; (7) reaction terminated by placing plugs into 0.5 M EDTA.

For digested DNA size-fractionation, agarose plugs were sliced using a sterile razor blade and placed in wells of a 1% agarose low-melting-point gel. DNA was size-separated with pulsed-field gel electrophoresis (PFGE, CHEF-DRII apparatus, Bio-Rad, Hercules) using run conditions of 6 V cm⁻¹ over 18 h with 10 to 60 s ramped switch-times. Separate molecular weight marker (Promega lambda ladder) and sample lanes were used for sizing without ethidium bromide staining and UV exposure of DNA for cloning.

DNA was electroeluted from the agarose gel using dialysis tubing. The gel area corresponding to the 100 to 125 kb fraction was excised and equilibrated in 1× tris-

ethylenediaminetetraacetic acid (TE) twice for 1 h at 4°C, then sliced into thin sections which were placed evenly into prepared dialysis tubing together with 600 µl 1× TE. DNA was electroeluted at 3 V cm⁻¹ for 3 h, at which time the current was reversed for 30 s to facilitate DNA removal from the dialysis tubing. The dialysis bag was then dialyzed twice for 1 h in deionized water at 4°C.

BAC cloning and arraying. We ligated 40 ng DNA into 50 ng *Hind*III-digested pCC1BAC cloning vector (Epicentre) according to the manufacturer's recommendations. This ligation mixture was incubated overnight at 16°C; the ligase was then inactivated at 55°C for 10 min. The ligation reaction was drop-dialyzed for 2 h using 0.025 µm VSWP filters (Millipore) on 0.5× TE. Then 20 µl of ligation reaction was electroporated into 20 µl Transformax EPI300 electrocompetent *E. coli* cells (Epicentre) using a Bio-Rad MicroPulser at 1.45 kV and immediately taken up in 0.5 ml iced SOC media; cells were allowed to recover at 37°C for 1 h. Transformants were stored in 25% glycerol at 80°C. Clones were sent to Benaroya Research Institute (Seattle) for plating and arraying onto positively-charged nylon membranes using a robotic colony-picking tool. Replicate membranes each contained 3072 clones, arrayed in a grid pattern corresponding to the clone position on one of eight 384-well plates.

Clone screening. Radiolabeled RNA riboprobes for *rbcL* were created as previously described (Paul et al. 1999). In short, probes were generated from Forms IA, IB, ID and II (*Rhodobacter sphaeroides*) *rbcL/cbbL/cbbM* fragments, along with archaeal *rbcL* fragments from *Methanococcus jannaschii*, *Methanosarcina acetivorans* C2A, and *Archaeoglobus fulgidus*, by *in vitro* transcription (Riboprobe Combination System SP6/T7, Promega) according to the manufacturer's instructions. Probes were labeled with ³⁵S-UTP (Amersham Biosciences). Hybridization and washing of filters was also as previously described (Paul et al. 1999). Hybridization was detected by autoradiography on x-ray film. Only *rbcL*-containing clones were isolated. Clones hybridizing to the probe were verified to contain the *rbcL* gene by PCR and colony-lifting and re-probing. Primers used for screening were Form ID primers 5'-GATGATGARAAYATTAATC forward, 5'-ATTTGDCCACAGTGDATACCA reverse; Form IA primers 5'-CTGAGIGGIAARAACACTACGG forward, 5'-GGCA-TRTGCCANACRTGRAT reverse.

Plasmid recovery and sequencing. Plasmids from desired clones were recovered by standard, alkaline-lysis plasmid-preparation procedures (omitting phenol-chloroform extraction) and 100% isopropanol precipitation followed by 70% ethanol washing (Sambrook & Russell 2001). BAC plasmids were sized after *Not*I digestion (Promega, following the manufacturer's instructions) using PFGE in a 1.25% PFGE-grade

agarose gel (BioWhittaker), with run conditions of 3.3 V cm⁻¹, 0.2 to 0.8 s ramped switch-times over 24 h at 14°C. Clones were sequenced at the US Department of Energy Joint Genome Institute (Walnut Creek, California). The accuracy of sequence assemblies was verified by comparing actual restriction fragments obtained with *Not*I digestion and PFGE separation to those predicted by restriction analysis performed *in silico* with sequence-analysis software (Kodon 2.01, Applied Maths). Kodon was used to perform open reading-frame analysis and record annotations. Genes in the sequence assemblies were manually annotated by determining the closest match to predicted open reading frames (ORFs) by BLAST searches (www.ncbi.nlm.nih.gov/BLAST/). For each putative coding region, the organism with the closest homologue and percent similarity (amino acid) was noted. These BAC sequences appear in Genbank under accession nos. DQ284919–DQ284920 and DQ325538–DQ325543.

RESULTS

We detected 8 unique BAC clones that hybridized with the Form IA *rbcL* probe; no other RubisCO genes were detected. These 8 clones also amplified with the Form IA *rbcL* primer set. A nucleotide identity matrix of the BAC sequences along with segments from *Synechococcus* sp. strain WH 8102 and 2 *Prochlorococcus marinus* strains is shown in Fig. 1. A total of 350 kb of genomic sequence from the environment was obtained and 364 presumptive open reading frames were identified.

All BAC genome fragments came from marine α -*Synechococcus*, based on *rbcL* nucleotide alignment (Fig. 2). Several clones were closely related based on *rbcL* sequence information; 2 pairs (3M9 + 4P21, and 5B2 + 4N23) shared 99% nucleotide identity and had the same gene content and order over their common lengths; 1 of each pair was shorter, with the end of each shorter sequence falling exactly on a *Hind*III restriction site on the longer one (*Hind*III was the enzyme originally used to digest community DNA for BAC library construction). Clones 3O6 and 4O4 are very closely related based on *rbcL* sequence and have 97% identity overall, but 3O6 has 2 genes inserted that are not present on 4O4.

Gene content and order was generally conserved among the BAC sequences in relation to the genome of *Synechococcus* sp. strain WH 8102, particularly in the genomic region surrounding *rbcL*. Fig. 3 shows the annotation of genes immediately surrounding *rbcL* of 3 BAC clones, WH 8102 and *Prochlorococcus marinus* strains MED4 and MIT 9313. The BAC clones not shown have the same order of genes in this region as WH 8102, except for BAC 4N23 (same as 5B2). This

| | | | | | | | | | | | | |
|------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| <i>Synech. sp.</i> WH 8102 | 100 | 49 | 79 | 77 | 68 | 94 | 55 | 47 | 81 | 18 | 6.5 | 76.0 |
| BAC 3M9 | 49 | 100 | 62 | 52 | 65 | 63 | 99 | 44 | 62 | 19 | 5.8 | 72.7 |
| BAC 3O12 | 79 | 62 | 100 | 80 | 68 | 76 | 56 | 64 | 84 | 23 | 8.0 | 40.5 |
| BAC 3O6 | 77 | 52 | 80 | 100 | 70 | 97 | 55 | 56 | 81 | 19 | 6.1 | 51.1 |
| BAC 4N23 | 68 | 65 | 68 | 70 | 100 | 67 | 65 | 99 | 64 | 23 | 12 | 19.6 |
| BAC 4O4 | 94 | 63 | 76 | 97 | 67 | 100 | 62 | 62 | 77 | 21 | 6.7 | 34.9 |
| BAC 4P21 | 55 | 99 | 56 | 55 | 65 | 62 | 100 | 52 | 61 | 19 | 6.1 | 43.5 |
| BAC 5B2 | 47 | 44 | 64 | 56 | 99 | 62 | 52 | 100 | 61 | 18 | 7.1 | 56.8 |
| BAC 5D20 | 81 | 62 | 84 | 81 | 64 | 77 | 61 | 61 | 100 | 25 | 10 | 35.5 |
| <i>Prochl. str.</i> MIT 9313 | 18 | 19 | 23 | 19 | 23 | 21 | 19 | 18 | 25 | 100 | 15 | 48.4 |
| <i>Prochl. str.</i> MED4 | 6.5 | 5.8 | 8.0 | 6.1 | 12 | 6.7 | 6.1 | 7.1 | 10 | 15 | 100 | 41.4 |

Fig. 1. Similarity matrix showing percent nucleotide identity along common length of each aligned sequence pair. Sizes of fragments/bacterial artificial chromosome (BAC) sequences compared are shown on the right. Published genome fragments were trimmed such that they are approximately centered on *rbcL*. *Synech.*: *Synechococcus*; *Prochl.*: *Prochlorococcus*

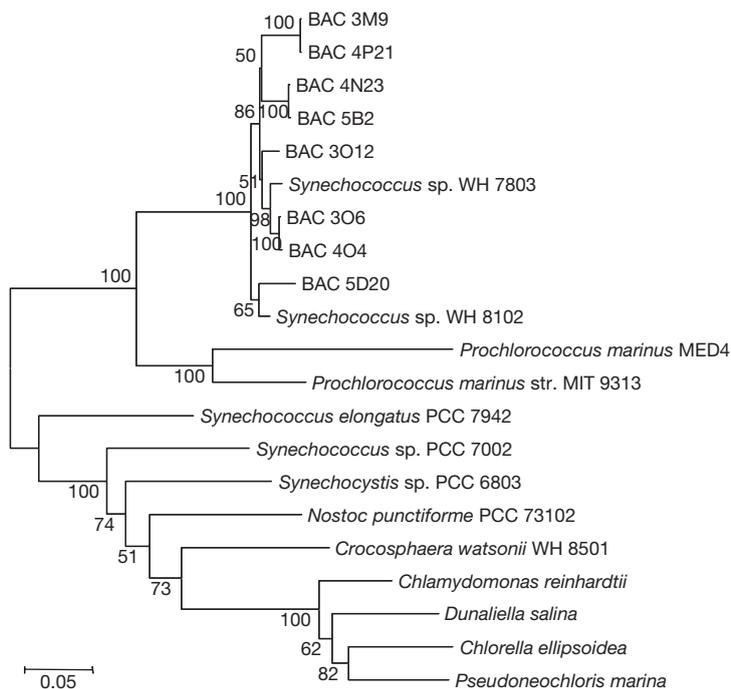
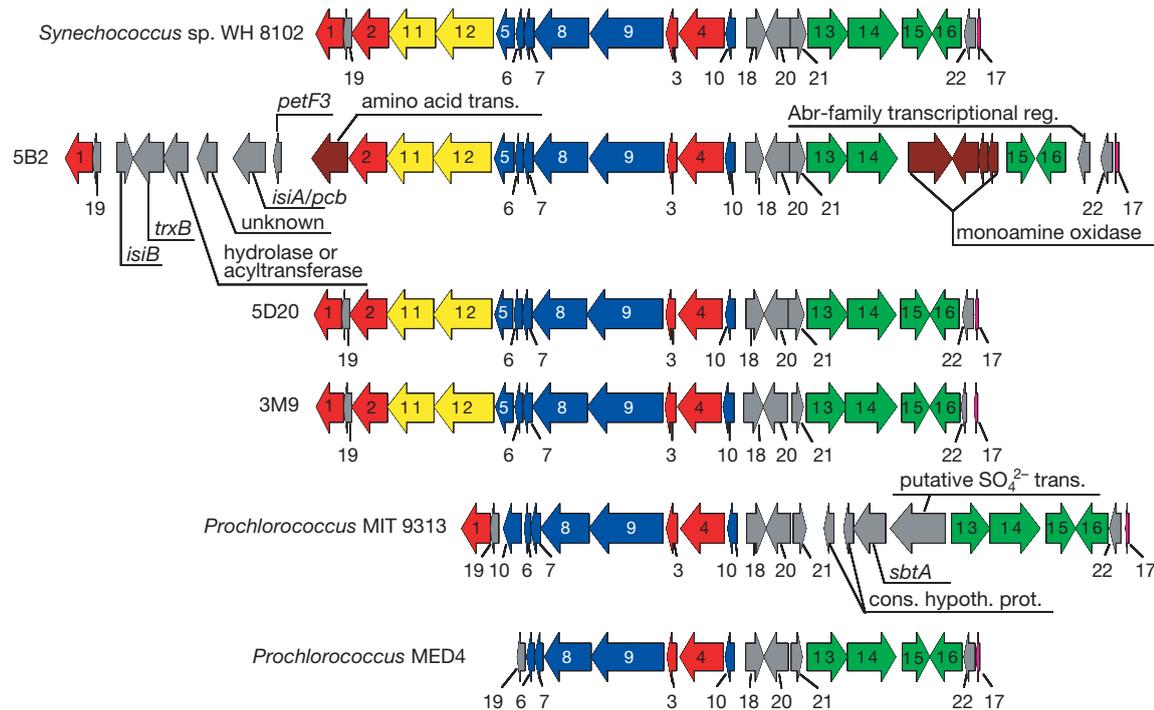


Fig. 2. Neighbor-joining tree showing taxonomic affiliation of bacterial artificial chromosome (BAC) clones with marine α -*Synechococcus* spp., based on *rbcL* nucleic acid sequences, except for *Dunaliella salina* and *Pseudoneochloris marina* which were aligned using an approx. 1350 bp fragment of *rbcL*. Tree constructed using a Kimura 2-parameter model with 1000 bootstrap repetitions on MEGA Version 2.1 (Kumar et al. 2001)

order was interrupted in 5B2 and 4N23 by groups of genes not present on the other BAC sequences. Amino acid sequence-similarity to WH 8102 in this region was high; for the 10 genes coding carbon-fixation structural or functional proteins shown in Fig. 3, homology to respective loci in WH 8102 for the 8 BAC clones ranged from 89.6 to 97.9% identity.

In *Prochlorococcus marinus*, the synteny of genes near *rbcL* is partially conserved in comparison to the BAC clones and *Synechococcus* sp. strain WH 8102. *P. marinus* strain MIT 9313 lacks the CO₂ hydration protein gene *chpX* and has several genes present within this series not found on WH 8102 or the BAC clones, including conserved hypothetical proteins and putative bicarbonate (*sbtA*) and sulfate transporters. *P. marinus* strain MED4 lacks both *chpX* and *cbbX*, and does not have 2 copies of the carboxysome shell peptide *csoS1* or homologous *ccmK1* gene. Aside from these differences, however, the sequence of genes from the carboxysome Subunit B gene to conserved hypothetical protein PMT1209 in MIT 9313, PMM0546 in MED4, and SYNW 1722 in WH 8102 (and its homologues in the BAC clones) is conserved.

A total of 22 genes with closest homology to cyanobacteria other than *Synechococcus* sp. WH 8102 were identified from the BAC clones (Table 1). Of the 22 identified, 20 do not have homologues in the WH 8102 genome; 2 of the BACs contained divergent regions containing several putative genes not found near *rbcL* on WH 8102 (Fig. 4). On one end of BAC 3M9 is a series of ORFs that most closely match a series of hypothetical protein genes in *Prochlorococcus marinus* strain MIT 9313, but are about 1 Mb away from *rbcL* and have only 15 to 35% similarity at the amino acid level. This gene sequence in MIT 9313 is adjacent to *lepA* (guanosine triphosphate-binding protein), which is found near *rbcL* on *Synechococcus* sp. WH 8102 and BAC 3M9. The other divergent region on BAC 3M9 is a series of ORFs with closest



| No. | Gene | Putative product | No. | Gene | Putative product |
|---|--------------------|--|---------------------------------------|-------------|---|
| Carbon fixation genes | | | Chlorophyll biosynthesis genes | | |
| 1 | <i>cbbX</i> | Rubis CO expression protein | 13 | <i>chlN</i> | light-independent protochlorophyllide reductase subunit |
| 2 | <i>chpX</i> | CO ₂ hydration protein | 14 | <i>chlB</i> | light-independent protochlorophyllide reductase subunit |
| 3 | <i>rbcs</i> | Rubis CO small subunit | 15 | <i>chlL</i> | protochlorophyllide reductase iron-sulfur ATP-binding protein |
| 4 | <i>rbcl</i> | Rubis CO large subunit | 16 | <i>pcr</i> | light dependent protochlorophyllide oxido-reductase |
| CO₂ concentrating/carboxysome protein genes | | | Photosystem protein genes | | |
| 5 | <i>ccmK2</i> | CO ₂ concentrating mechanism protein | 17 | <i>psaM</i> | photosystem I reaction centre subunit XII |
| 6 | | carboxysome peptide B | Other/unknown function | | |
| 7 | | carboxysome peptide A | 18 | <i>ham1</i> | HAM 1 family protein |
| 8 | <i>csoS3</i> | carboxysome shell polypeptide | 19 | | Pterin-4a-carbinolamine dehydratase [Coenzyme metabolism] |
| 9 | <i>csoS2</i> | carboxysome shell polypeptide | 20 | | conserved hypothetical protein |
| 10 | <i>csoS1/ccmK1</i> | CO ₂ concentrating mechanism-structural protein | 21 | | conserved hypothetical protein |
| NADH dehydrogenase genes | | | 22 | | conserved hypothetical protein |
| 11 | <i>ndhD2</i> | NADH dehydrogenase I chain 4 (or M) | | | |
| 12 | <i>ndhF1</i> | NADH dehydrogenase I chain 5 (or L) | | | |

Fig. 3. Synteny of genes near *rbcl* for 3 bacterial artificial chromosome (BAC) sequences, *Synechococcus* sp. strain WH 8102, and 2 *Prochlorococcus* genomes. Genes are color-coded based on related functions. Red: CO₂ fixation; blue: carboxysome; yellow: NADH dehydrogenase; green: chlorophyll biosynthesis; brown: amine/amino acid transport or metabolism; pink: photosystem (No. 17). Gray: other or unknown function

homology to the respective genes in *Synechococcus* sp. WH 8102, with gene order matching that of WH 8102, but from a genomic region about 700 kb away from *rbcl*.

There are several deviations in genomic organization from *Synechococcus* sp. WH 8102 on BAC 5B2 (Fig. 4). These include putative genes for an amino acid transporter and 2 monoamine oxidases not found in the WH 8102 genome, with closest homology to genes from *Crocospaera watsonii* and *Synechocystis* sp. PCC 6803 respectively (Table 1). Of the 2

monoamine oxidases, one is a single ORF, the other (in the reverse direction) consists of 3 ORFs (i.e. include 2 stop codons) that together constitute a homologue. Other genes divergent from WH 8102 include several potentially involved in light-harvesting electron transfer (homologues of *isiB*, coding for an iron-stress-induced flavodoxin, the phylogenetically-ubiquitous thioredoxin reductase, a putative hydrolyase or acyltransferase, and a chlorophyll-binding protein similar to iron-stress antenna-structure proteins such as *IsiA* and *Pcb* in cyanobacteria).

Table 1. Putative genes from bacterial artificial chromosome (BAC) clones with closest homology to those of cyanobacteria other than *Synechococcus* sp. strain WH 8102. *P. marinus*; *Prochlorococcus marinus*; AA: amino acid

| BAC | Homologous gene | Homologous gene product | Organism of closest homologue | AA similarity to protein (%) | Homologue in WH 8102? |
|------|-------------------------|---|-------------------------------------|------------------------------|-----------------------|
| 3M9 | | Possible pilin | <i>P. marinus</i> str. MIT 9313 | 35 | no |
| 3M9 | | Hypothetical protein PMT0261 | <i>P. marinus</i> str. MIT 9313 | 30 | no |
| 3M9 | | Hypothetical protein PMT0260 | <i>P. marinus</i> str. MIT 9313 | 15 | no |
| 3M9 | | Hypothetical protein PMT0259 | <i>P. marinus</i> str. MIT 9313 | 19 | no |
| 3M9 | | Hypothetical protein PMT0258 | <i>P. marinus</i> str. MIT 9313 | 30 | no |
| 3M9 | | Small mechanosensitive ion channel, MscS family | <i>P. marinus</i> str. MIT 9313 | 52 | yes |
| 3M9 | | Partial hypothetical protein PMT1393 | <i>P. marinus</i> str. MIT 9313 | 9 | no |
| 3O12 | | Transposase and inactivated derivatives, IS5 family | <i>Nostoc punctiforme</i> | 44 ^a | no |
| 3O12 | | Hypothetical protein all2909 | <i>Nostoc</i> sp. PCC 7120 | 30 | no |
| 3O6 | | Ca ²⁺ -binding protein (EF-Hand superfamily) | <i>N. punctiforme</i> | 39 | no |
| 4N23 | <i>Npun2019</i> | Monoamine oxidase | <i>Synechocystis</i> sp. PCC 6803 | 34 | no |
| 5B2 | <i>Npun2019</i> | Ca ²⁺ -binding protein (EF-Hand superfamily) | <i>N. punctiforme</i> | 39 | no |
| 5B2 | <i>isiB</i> | Partial hypothetical protein PMT0982 | <i>P. marinus</i> str. MIT 9313 | 9 | no |
| 5B2 | <i>trxB</i> | Flavodoxin | <i>P. marinus</i> SS120 | 53 | no |
| 5B2 | | Thioredoxin reductase | <i>P. marinus</i> SS120 and MED 4 | 62 | yes |
| 5B2 | | Predicted hydrolase or acyltransferase | <i>Crocospaera watsonii</i> WH 8501 | 32 | no |
| 5B2 | | Unknown protein sll1549 ^b | <i>Synechocystis</i> sp. PCC 6803 | 33 | no |
| 5B2 | <i>pcbD^c</i> | Chlorophyll <i>a/b</i> binding light harvesting protein | <i>P. marinus</i> SS120 | 62 | no |
| 5B2 | | Amino acid transporter | <i>C. watsonii</i> WH 8501 | 12 | no ^d |
| 5B2 | | Monoamine oxidase | <i>Synechocystis</i> sp. PCC 6803 | 37 | no |
| 5B2 | | Monoamine oxidase ^e | <i>Synechocystis</i> sp. PCC 6803 | 31 | no |
| 5B2 | | Conserved bacterial protein all2909 | <i>Nostoc</i> sp. PCC 7120 | 35 | no |

^a3 successive ORFs on BAC 3O12 combine for homology to single *Nostoc* gene

^bSome internal similarity to conserved domains of phosphoglycerate mutase family, fructose-2,6-bisphosphatase and phosphohistidine phosphatase 6 A of various bacteria

^cNear-equal homology to several other *Prochlorococcus* (*pcb*) and *Synechococcus/Synechocystis* (*isiA*) chlorophyll-binding protein genes (primarily iron-stress-induced)

^dNo homologue found to this gene by amino acid BLAST searching, but ABC family-type amino acid membrane-transport protein genes have been reported in WH 8102 (Palenik et al. 2003)

^e3 successive ORFs on BAC 5B2 combine for homology to single *Synechocystis* gene

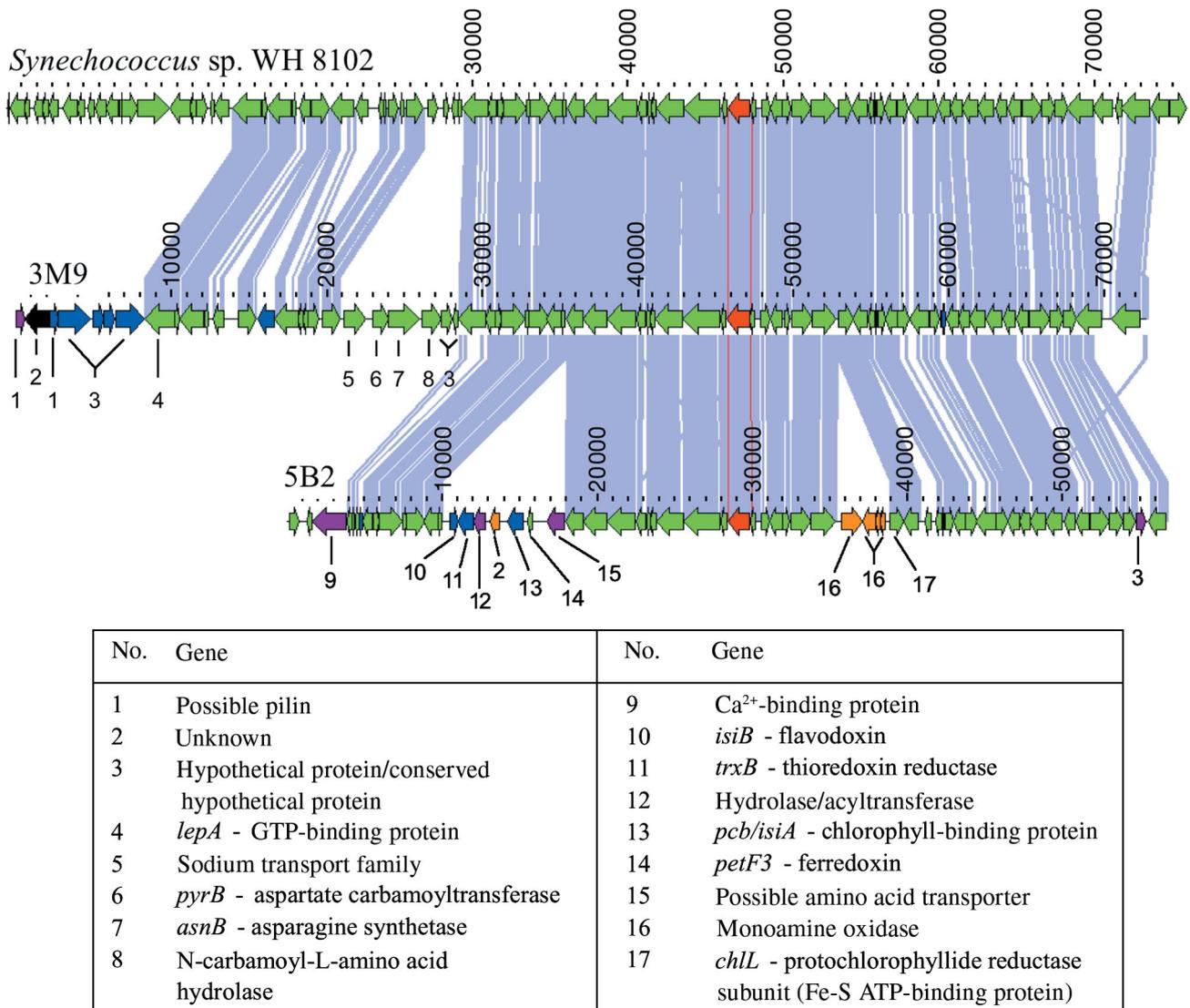


Fig. 4. Alignments showing homology of BAC 3M9 to genomic region surrounding *rbcL* on *Synechococcus* sp. strain WH 8102 and homology of BAC 5B2 to BAC 3M9. *rbcL* is colored red for reference. Blue-shaded areas indicate homology features of at least 80 nucleotides having at least 60% identity. Presumptive genes on BAC clones are color-coded, showing organisms to which they have closest homology. Green: *Synechococcus* sp. WH 8102; blue: *Prochlorococcus*; orange: *Synechocystis*; purple: other cyanobacteria including *Nostoc* and *Trichodesmium*; black: unknown. BAC: bacterial artificial chromosome

DISCUSSION

A BAC library constructed from an oligotrophic environment in the Gulf of Mexico was screened for *rbcL*-containing genomic fragments in an effort to understand the 'genomic environment' of uncultivated phytoplankton occurring there. All *rbcL*-containing BACs were from phycoerythrin-containing *Synechococcus* spp. The number of clones containing *Synechococcus rbcL* in our BAC library is in agreement with estimates of the abundance of *Synechococcus* within the sample. Flow-cytometry of this sample yielded 5.7×10^4 cells ml⁻¹ (Lisa Campbell, Texas A&M University,

pers. comm.), and assuming a DNA content of $4 \mu\text{g l}^{-1}$ (Paul et al. 1985), the *Synechococcus* genomic fraction would be 3.8% of the microbial DNA present. Because 1.85% of a BAC library comprised of 50 kb clones from a pure culture of *Synechococcus* would be expected to contain the *rbcL* gene, then 0.07% of the clones of a 3200 Gulf of Mexico metagenomic library (or 2.25) would be positive for *Synechococcus rbcL*. In reality, we found 8 such clones, and if the DNA content of the waters was actually $1 \mu\text{g}$, this would precisely agree with our *rbcL* clone yield. However, even more *Prochlorococcus* cells were present in this sample, as well as some picoeucaryotes, and no clones with these

types of *rbcL* were obtained. It may be that these cells were not as efficiently cloned, through an as yet unknown selective process in the BAC cloning procedure. Prior work has shown that our filtration process is 70 to 90 % efficient for harvesting phytoplankton and bacterioplankton (Jiang et al. 1992). It may be that certain cell types are resistant to the lysis procedure performed or have DNA less susceptible to *Hind*III cleavage.

Analysis of the 8 α -*Synechococcus* genomic fragments revealed a general conservation of genes near RubisCO with those of *Synechococcus* sp. strain WH 8102, as well as several divergent regions. We found that 2 clones contained clusters of genes divergent from WH 8102 (3M9 and 5B2: Fig. 4); 10 % (3M9) and 19 % (5B2) of the ORFs on these BACs were not found in WH 8102. The remaining clones contained at most 1 or 2 genes not found on WH 8102 (>94 % of ORFs with the closest match to WH 8102). About $\frac{2}{3}$ of the open reading frames on the complete genome of *Synechococcus* sp. WH 8102 are shared with one or both *Prochlorococcus marinus* strains MIT 9313 or MED4 (Palenik et al. 2003, Rocap et al. 2003). A primary divergence in gene organization in the *rbcL* genomic environment among fully-sequenced marine picocyanobacteria is the presence of NADH dehydrogenases (i.e. genes *ndhD2* and *ndhF1*) in WH 8102, which are lacking in *Prochlorococcus* strains MED4 and MIT 9313 (Badger et al. 2002). This key contrast to *Prochlorococcus* organisms is conserved in our uncultivated *Synechococcus* as well.

The additional genes found within the CO₂-fixation gene cluster on BAC Clones 5B2 and 4N23 suggest that insertions resulting from horizontal gene transfer may have occurred in this conserved operonic region. Besides the large divergent areas on BAC Clones 3M9 and 5B2, other BACs show evidence of smaller insertions or deletions, such as the presence of calcium-binding protein genes on 3O6 and 5B2, whose closest homologue is found in the cyanobacterium *Nostoc punctiforme* (Table 1).

Of ecological significance is the finding that the BAC clones contain genes coding for proteins potentially important for organic N transport and metabolism (putative amino acid transport-system protein and monoamine oxidases on BAC 5B2 and 4N23). The putative amino acid transport-system protein is unrelated to previously described proteins identified on the genome of *Synechococcus* sp. strain WH 8102 (Palenik et al. 2003), and is most closely related to a gene found in *Crocospaera watsonii* WH 8501 (23 % amino acid identity over 50 % of the protein: Table 1).

The monoamine oxidase genes on BACs 5B2 and 4N23 are even more intriguing. Monoamine oxidases convert organic amines to ammonia and the organic C backbone while producing peroxide (Zubay et al.

1995). The closest homologue to these sequences is in *Synechocystis* sp. strain PCC 6803; monoamine oxidases are also found in the genomes of *Synechococcus elongatus* strains PCC 7942 and PCC 6301, all β -cyanobacteria. The sequences obtained from 5B2 and 4N23 also contain strong homology to flavin-containing amine oxidoreductases determined by a conserved domain database search with BLAST. Marine cyanobacteria have been shown to utilize organic nitrogen to fulfill their nutrient requirements (e.g. some *Prochlorococcus* spp. use amino acids), enabling efficient N recycling and possibly leading to their greater prevalence in highly oligotrophic waters (Paerl 1991, Zubkov et al. 2003). The identification of monoamine oxidase genes in environmental *Synechococcus* genomic fragments suggests that amines may potentially serve as a nitrogen source for marine cyanobacteria

A metagenomic sequence from the Sargasso Sea (Accession No. AACY01342157.1) shared 88 % amino acid identity to the last $\frac{1}{2}$ of the amine oxidase ORF on 5B2. However, an interesting aspect of this fragment is that it also contained the first part of an ORF with homology to the *chlL* gene of *Synechococcus* sp. strain WH 8102 (94 % amino acid identity to ChlL of WH 8102, from the start of ORF to the end of the fragment, the closest match in Genbank). *chlL* codes for a protochlorophyllide reductase subunit, part of the chlorophyll biosynthesis pathway, and is conserved among many photosynthetic bacteria. The gene arrangement on the Sargasso Sea fragment is similar to that on BAC 5B2 (Fig. 4) except for the absence of the reverse amine oxidase. From the high level of similarity between the *chlL* genes, it appears that the Sargasso Sea fragment was also derived from *Synechococcus* spp., further evidence that some marine α -*Synechococcus* spp. contain a monoamine oxidase gene, most likely in the *rbcL* genomic region.

Our finding of an *isiA/pcb*-type gene on BAC Clone 5B2 is noteworthy, as it is the first evidence of chlorophyll-binding antenna proteins among α -*Synechococcus*. The closest homologue of the chlorophyll binding protein gene in BAC 5B2 (*pcb/isiA* on Fig. 4) is the PcbD protein of *Prochlorococcus marinus* SS120 (CCMP1375) with 62 % amino acid identity, but significant homology also existed to several IsiA iron-stress-induced chlorophyll-binding proteins. The cladogram in Fig. 5 indicates that the *isiA/pcb* gene product from 5B2 is deeply rooted between the *Prochlorococcus* Pcb proteins and the IsiA proteins of β -cyanobacteria. These proteins form complexes with Photosystem I (PSI) and II (PSII) units to enhance photosynthetic efficiency under conditions of iron starvation, when PSI becomes depleted due to inability to form the integral iron-sulfur reaction centers (Bibby et al. 2001a,b,

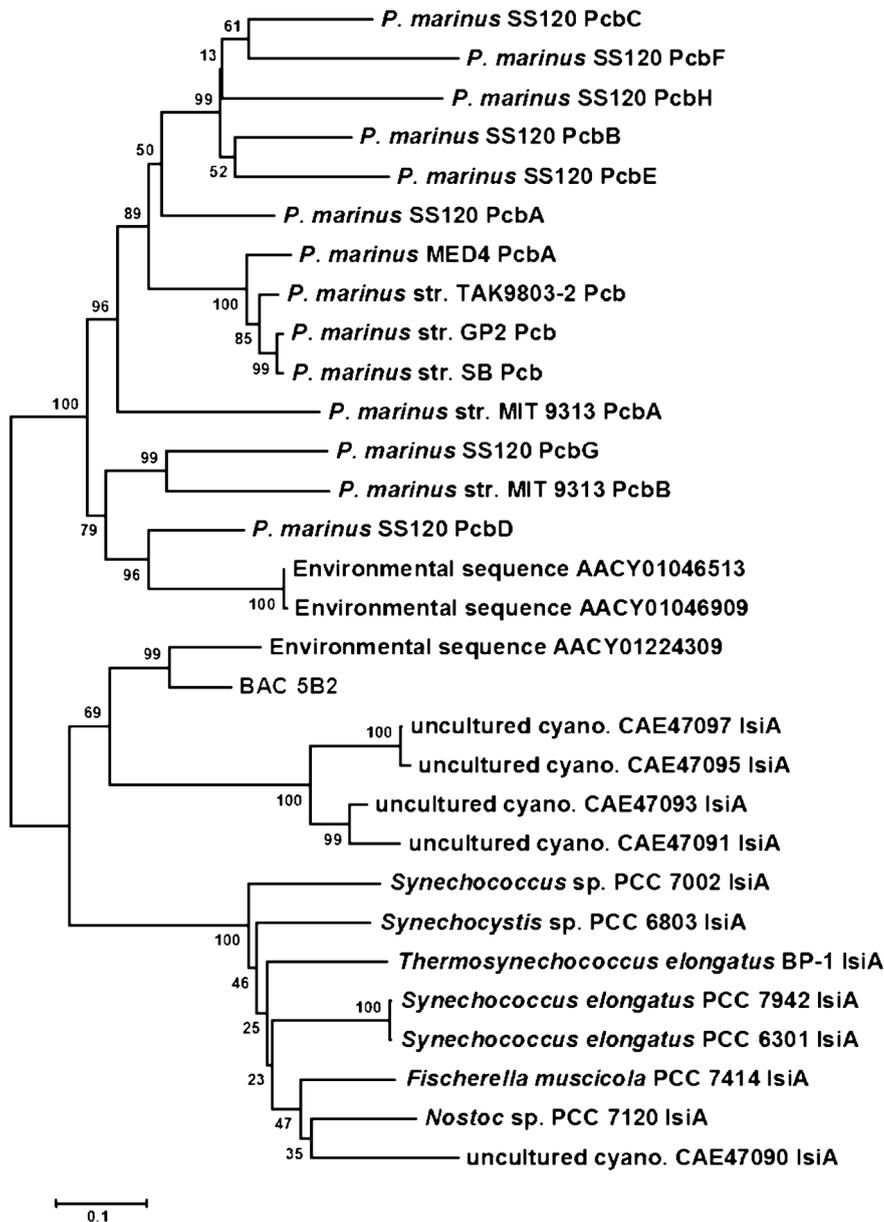


Fig. 5. Neighbor-joining tree from amino acid alignment of several IsiA and Pcb protein sequences from cyanobacteria and uncultivated sequences (available from GenBank). Uncultured cyanobacteria IsiA sequences were cloned by PCR using primers specific for *isiA* genes (Geiss et al. 2004); environmental sequences are from the Sargasso Sea metagenome (Venter et al. 2004). *P. marinus*: *Prochlorococcus marinus*. Numbers show scores of 1000 bootstrap repetitions. Cladogram constructed using Poisson correction method on MEGA Version 2.1 (Kumar et al. 2001)

Boekema et al. 2001, Yeremenko et al. 2004). In some *Prochlorococcus* spp., Pcb proteins when expressed constitutively serve to adapt the organism for enhanced photosynthetic light-harvesting ability in dimly-lit deeper waters (Bibby et al. 2003). In marine *Synechococcus* sp. strain PCC 7002 and several freshwater unicellular β -cyanobacteria, IsiA and flavodoxin are co-expressed in response to iron depletion (Lau-

denbach & Straus 1988, Leonhardt & Straus 1992, Vinnemeier et al. 1998). Flavodoxin is an alternate terminal electron-acceptor to ferredoxin under conditions of iron limitation. Given its proximity to the flavodoxin gene, the *isiA/pcb* gene in 5B2 may be part of an iron-stress operon with a function similar to that of IsiA proteins. Alternatively, it is possible that this gene may encode a constitutively-expressed antenna protein to enable enhanced light-harvesting by PSI or PSII, much as PcbA and PcbG of *P. marinus* strains MIT 9313 and SS120 respectively (Bibby et al. 2003). In such case, BAC 5B2 may represent a novel type of low-light-adapted marine *Synechococcus*. Coupled with the putative organic nitrogen-utilization genes, BAC 5B2 may represent a *Synechococcus* strain with considerably expanded metabolic capabilities.

The similarity of IsiA and Pcb proteins has been attributed to a common ancestor of *Prochlorococcus* and *Synechococcus*, despite the lack of a homologue in *Synechococcus* sp. strain WH 8102 (Bibby et al. 2003). However, these commonalities are more likely due to gene-transfer events, as the appearance of the divergent segment containing this gene on BAC 5B2 suggests. This segment contains thioredoxin reductase, ferredoxin and flavodoxin, and may be an electron-transfer gene 'cassette', providing a mechanism of adaptation via horizontal gene transfer. Since it appears within a highly conserved operonic region, it seems likely that this gene cluster occurred as an insertion rather than a deletion from a common ancestor. On BAC 3M9, the homology to *lepA* at

the end of the *Prochlorococcus* gene series may have enabled their movement by homologous recombination, perhaps resulting from a transformation event between *Prochlorococcus* and *Synechococcus*. It seems likely that horizontal gene-transfer events may introduce considerable genomic diversity into marine cyanobacteria populations. Thus, while highly critical gene clusters may be more tightly conserved, as for carbon fixation, exchange

of other functional genes between cyanobacterial clades may be a mechanism for oceanic niche-adaptation.

The functional diversity of marine picocyanobacteria is most certainly greater than the diversity revealed by key phylogenetic markers, either structural genes such as 16S rDNA sequences or functional genes such as *rbcL*. What may become more evident as diversity and ecological studies advance from the genetic to the genomic scale, is that microdiversity of cyanobacterial subtypes is enabled by frequent horizontal gene-transfer events. Indeed, *Synechococcus* spp. and *Prochlorococcus* photosynthesis genes, including a *petF* gene for ferredoxin (also part of the divergent gene cluster on BAC 5B2; Fig. 4), have been found on the genomes of cyanophages which infect them (Mann et al. 2003, Lindell et al. 2004). Clearly, genomic information inferred about entire groups of cyanobacteria based on sequencing of a few type-strains does not paint a complete picture of the ecological significance and diversity present in the various oceanic environments. Along with continued analyses of nutrient response and physiological aspects, environmental genomics will play an ever-larger role in understanding the ecology of oceanic microorganisms as they relate to biogeochemical cycles.

After submission of the revised version of this paper, additional *Synechococcus* genomes have been released on GenBank. *Synechococcus* sp. CC9605 contains a putative amine oxidase gene in the same relative genomic position as that on our BAC 5B2 and is about 84% similar. It is thus possible that *Synechococcus* sp. CC9605 could prove valuable for investigating the possible utilization of amines among marine *Synechococcus*.

Acknowledgements. This research was supported by the Office of Science (BER), US Department of Energy, Grant Nos. DE-FG02-97ER62452 and DE-FG02-97ER62454 and by the National Science Foundation. Sequencing was provided by Paul Richardson and the US Department of Energy Joint Genome Institute. We also acknowledge the support of the crew of the RV 'F. G. Walton Smith'.

LITERATURE CITED

- Badger MR, Hanson D, Price GD (2002) Evolution and diversity of CO₂ concentrating mechanisms in cyanobacteria. *Funct Plant Biol* 29:161–173
- Beja O, Aravind L, Koonin EV, Suzuki MT and 8 others (2000a) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289:1902–1906
- Beja O, Suzuki MT, Koonin EV, Aravind L and 8 others (2000b) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* 2:516–529
- Bibby TS, Nield J, Barber J (2001a) Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* 412:743–745
- Bibby TS, Nield J, Partensky F, Barber J (2001b) Oxyphotobacteria—antenna ring around photosystem I. *Nature* 413:590
- Bibby TS, Mary I, Nield J, Partensky F, Barber J (2003) Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* 424:1051–1054
- Boekema EJ, Hifney A, Yakushevska AE, Piotrowski M and 5 others (2001) A giant chlorophyll–protein complex induced by iron deficiency in cyanobacteria. *Nature* 412:745–748
- Button DK (2001) Isolation of oligobacteria. In: Paul JH (ed) *Marine microbiology*, Vol 30. Academic Press, New York, p 161–173
- Corredor JE, Wawrik B, Paul JH, Tran H, Kerkhof L, Lopez JM, Dieppa A, Cardenas O (2004) Geochemical rate-RNA integrated study: ribulose-1,5-bisphosphate carboxylase/oxygenase gene transcription and photosynthetic capacity of planktonic photoautotrophs. *Appl Environ Microbiol* 70:5459–5468
- Dumont MG, Murrell JC (2005) Stable isotope probing—linking microbial identity to function. *Nature Rev Microbiol* 3:499–504
- Finn MW, Tabita FR (2003) Synthesis of catalytically active form III ribulose 1,5-bisphosphate carboxylase/oxygenase in archaea. *J Bacteriol* 185:3049–3059
- Geiss U, Selig U, Schumann R, Steinbruch R, Bastrop R, Hagemann M, Schoor A (2004) Investigations on cyanobacterial diversity in a shallow estuary (Southern Baltic Sea) including genes relevant to salinity resistance and iron starvation acclimation. *Environ Microbiol* 6:377–387
- Hanson TE, Tabita FR (2001) A ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. *Proc Natl Acad Sci USA* 98:4397–4402
- Hanson TE, Tabita FR (2003) Insights into the stress response and sulfur metabolism revealed by proteome analysis of a *Chlorobium tepidum* mutant lacking the Rubisco-like protein. *Photosynth Res* 78:231–248
- Jiang SC, Thurmond JM, Pichard SL, Paul JH (1992) Concentration of microbial populations from aquatic environments by vortex flow filtration. *Mar Ecol Prog Ser* 80:101–107
- Klenk HP, Clayton RA, Tomb JF, White O and 47 others (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390:364–370
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Laudenbach DE, Straus NA (1988) Characterization of a cyanobacterial iron stress-induced gene similar to psbC. *J Bacteriol* 170:5018–5026
- Leonhardt K, Straus NA (1992) An iron stress operon involved in photosynthetic electron transport in the marine cyanobacterium *Synechococcus* sp. PCC 7002. *J Gen Microbiol* 138:1613–1621
- Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA* 101:11013–11018
- Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003) Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* 424:741
- Paelr HW (1991) Ecophysiological and trophic implications of light-stimulated amino-acid utilization in marine picoplankton. *Appl Environ Microbiol* 57:473–479

- Palenik B, Brahamsha B, Larimer FW, Land M and 11 others (2003) The genome of a motile marine *Synechococcus*. *Nature* 424:1037–1042
- Paul JH, Jeffrey WH, Deflaun M (1985) Particulate DNA in sub-tropical oceanic and estuarine planktonic environments. *Mar Biol* 90:95–101
- Paul JH, Pichard SL, Kang JB, Watson GMF, Tabita FR (1999) Evidence for a clade-specific temporal and spatial separation in ribulose biphosphate carboxylase gene expression in phytoplankton populations off Cape Hatteras and Bermuda. *Limnol Oceanogr* 44:12–23
- Raven JA (1995) Phycological reviews. 15. Photosynthetic and nonphotosynthetic roles of carbonic-anhydrase in algae and cyanobacteria. *Phycologia* 34:93–101
- Robertson BR, Button DK, Koch AL (1998) Determination of the biomasses of small bacteria at low concentrations in a mixture of species with forward light scatter measurements by flow cytometry. *Appl Environ Microbiol* 64:3900–3909
- Rocap G, Larimer FW, Lamerdin J, Malfatti S and 20 others (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042–1047
- Rondon MR, August PR, Bettermann AD, Brady SF and 12 others (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* 66:2541–2547
- Sabehi G, Massana R, Bielawski JP, Rosenberg M, DeLong EF, Beja O (2003) Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ Microbiol* 5:842–849
- Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual*, Vol 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Schleper C, Swanson RV, Mathur EJ, DeLong EF (1997) Characterization of a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. *J Bacteriol* 179:7803–7811
- Schmidt TM, DeLong EF, Pace NR (1991) Analysis of a marine picoplankton community by 16S ribosomal-RNA gene cloning and sequencing. *J Bacteriol* 173:4371–4378
- Sieracki ME, Viles CL (1992) Distributions and fluorochrome staining properties of submicrometer particles and bacteria in the North Atlantic. *Deep-Sea Res A* 39:1919–1929
- Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* 178:591–599
- Tabita FR (1988) Molecular and cellular regulation of autotrophic carbon dioxide fixation in microorganisms. *Microbiol Rev* 52:155–189
- Tabita FR (1995) The biochemistry and metabolic regulation of carbon metabolism and CO₂ fixation in purple bacteria. In: Blankenship RE, Madigan MT, Bauer CE (eds) *Anoxygenic photosynthetic bacteria*. Kluwer Academic Press, Dordrecht, p 885–914
- Tabita FR (1999) Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth Res* 60:1–28
- Tabita FR, Hanson TE (2004) Anoxygenic photosynthetic bacteria. In: Fraser CM, Nelson KE, Read TD (eds) *Microbial genomics*. Humana Press, Totowa, NJ, p 225–243
- Venter JC, Remington K, Heidelberg JF, Halpern AL and 19 others (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–74
- Vinnemeier J, Kunert A, Hagemann M (1998) Transcriptional analysis of the isiAB operon in salt-stressed cells of the cyanobacterium *Synechocystis* sp. PCC 6803. *FEMS Microbiol Lett* 169:323–330
- Watson GMF, Yu JP, Tabita FR (1999) Unusual ribulose 1,5-bisphosphate carboxylase/oxygenase of anoxic Archaea. *J Bacteriol* 181:1569–1575
- Wawrik B, Paul JH (2004) Phytoplankton community structure and productivity along the axis of the Mississippi River plume in oligotrophic Gulf of Mexico waters. *Aquat Microb Ecol* 35:185–196
- Wawrik B, Paul JH, Campbell L, Griffin D, Houchin L, Fuentes-Ortega A, Muller-Karger F (2003) Vertical structure of the phytoplankton community associated with a coastal plume in the Gulf of Mexico. *Mar Ecol Prog Ser* 251:87–101
- Yeremenko N, Kouril R, Ihalainen JA, D'Haene S and 8 others (2004) Supramolecular organization and dual function of the IsiA chlorophyll-binding protein in cyanobacteria. *Biochemistry* 43:10308–10313
- Zubay GL, Parson WW, Vance DE (1995) *Principles of biochemistry*, Vol 1. Wm C Brown, Dubuque, IA
- Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R (2003) High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl Environ Microbiol* 69:1299–1304

Editorial responsibility: Otto Kinne (Editor-in-Chief), Oldendorf/Luhe, Germany

*Submitted: March 2, 2005; Accepted: November 28, 2005
Proofs received from author(s): June 13, 2006*