



## UvA-DARE (Digital Academic Repository)

### Shades of red and green : the colorful diversity and ecology of picocyanobacteria in the Baltic Sea

Haverkamp, T.H.A.

**Publication date**  
2008

[Link to publication](#)

#### **Citation for published version (APA):**

Haverkamp, T. H. A. (2008). *Shades of red and green : the colorful diversity and ecology of picocyanobacteria in the Baltic Sea*. [Thesis, externally prepared, Universiteit van Amsterdam]. Netherlands Institute of Ecology (NIOO) - Royal Netherlands Academy of Arts and Sciences.

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

# Chapter 3

## *Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons*

Thomas Haverkamp<sup>1</sup>  
Silvia G. Acinas<sup>1†</sup>  
Marije Doeleman<sup>1</sup>  
Maayke Stomp<sup>2</sup>  
Jef Huisman<sup>1,2</sup>  
Lucas J. Stal<sup>1,2\*</sup>

<sup>1</sup>Department of Marine Microbiology, Netherlands Institute of Ecology, NIOO-KNAW,  
P.O.Box 140, 4400 AC Yerseke, The Netherlands

<sup>2</sup>Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam,  
Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands

<sup>†</sup> Present address: Institut de Ciències del Mar, CMIMA, 08003, Barcelona, Spain

\*For correspondence. E-mail [L.Stal@nioo.knaw.nl](mailto:L.Stal@nioo.knaw.nl); Tel. (+31) 113 577300; Fax (+31) 113 573616.



Published in: Environmental Microbiology. Volume 10 pages: 174-188. Year: 2008

Running title: Phylogeny of red and green picocyanobacteria.

## Summary

*Picocyanobacteria of the genus Synechococcus span a range of different colours, from red strains rich in phycoerythrin (PE) to green strains rich in phycocyanin (PC). Here, we show that coexistence of red and green picocyanobacteria in the Baltic Sea is widespread. The diversity and phylogeny of red and green picocyanobacteria was analysed using three different genes: 16S rRNA-ITS, the cpeBA operon of the red PE pigment, and the cpcBA operon of the green PC pigment. Sequencing of 209 clones showed that Baltic Sea picocyanobacteria exhibit high levels of microdiversity. The partial nucleotide sequences of the cpcBA and cpeBA operons from the clone libraries of the Baltic Sea revealed two distinct phylogenetic clades: one clade containing mainly sequences from cultured PC-rich picocyanobacteria, while the other contains only sequences from cultivated PE-rich strains. A third clade of phycourobilin (PUB) containing strains of PE-rich Synechococcus spp. did not contain sequences from the Baltic Sea clone libraries. These findings differ from previously published phylogenies based on 16S rRNA gene analysis. Our data suggest that, in terms of their pigmentation, Synechococcus spp. represent three different lineages occupying different ecological niches in the underwater light spectrum. Strains from different lineages can coexist in light environments that overlap with their light absorption spectra.*

## Introduction

Picocyanobacteria of the *Synechococcus* group span a range of different colours, depending on their pigment composition (Wood, 1985; Olson *et al.*, 1990; Pick, 1991; Vörös *et al.*, 1998; Stomp *et al.*, 2007). Picocyanobacteria with high concentrations of the pigment phycoerythrin (PE) absorb green light effectively, and have a red appearance. Picocyanobacteria with high concentrations of phycocyanin (PC) absorb red light effectively, and have a blue-green colour. Recent competition models and laboratory experiments showed that red picocyanobacteria win the competition in green light, green picocyanobacteria win in red light, while red and green picocyanobacteria can coexist in white light by partitioning of the light spectrum (Stomp *et al.*, 2004). This matches their distribution patterns. Red picocyanobacteria are dominant components of the *Synechococcus* group in open ocean waters (Li *et al.*, 1983; Platt *et al.*, 1983; Campbell and Carpenter, 1987; Campbell and Vaulot, 1993), where green and particularly blue light penetrate deeply into the water column. Moreover, red picocyanobacteria can have two different bilin pigments known as phycoerythrobilin (PEB) and phycourobilin (PUB), which both bind to the apoprotein phycoerythrin. The absorption peak of PUB is shifted slightly further to the blue part of the spectrum, and picocyanobacteria with a high PUB/PEB ratio are typically dominant in oligotrophic regions of the oceans where blue light prevails (Olson *et al.*, 1990; Wood *et al.*, 1998; Toledo *et al.*, 1999). In addition, some strains are able to modify their pigmentation through the synthesis of PE with two alternative chromophores, PEB and PUB (Type IV CA; Everroad *et al.*, 2006). Green picocyanobacteria dominate in turbid waters, where red light prevails (Stomp *et al.*, 2007). Coexistence of red and green picocyanobacteria can be found in waters of intermediate colouration, including coastal seas and many freshwater lakes (Pick, 1991; Vörös *et al.*, 1998; Murrell and Lores, 2004; Katano *et al.*, 2005; Mózes *et al.*, 2006; Stomp *et al.*, 2007).

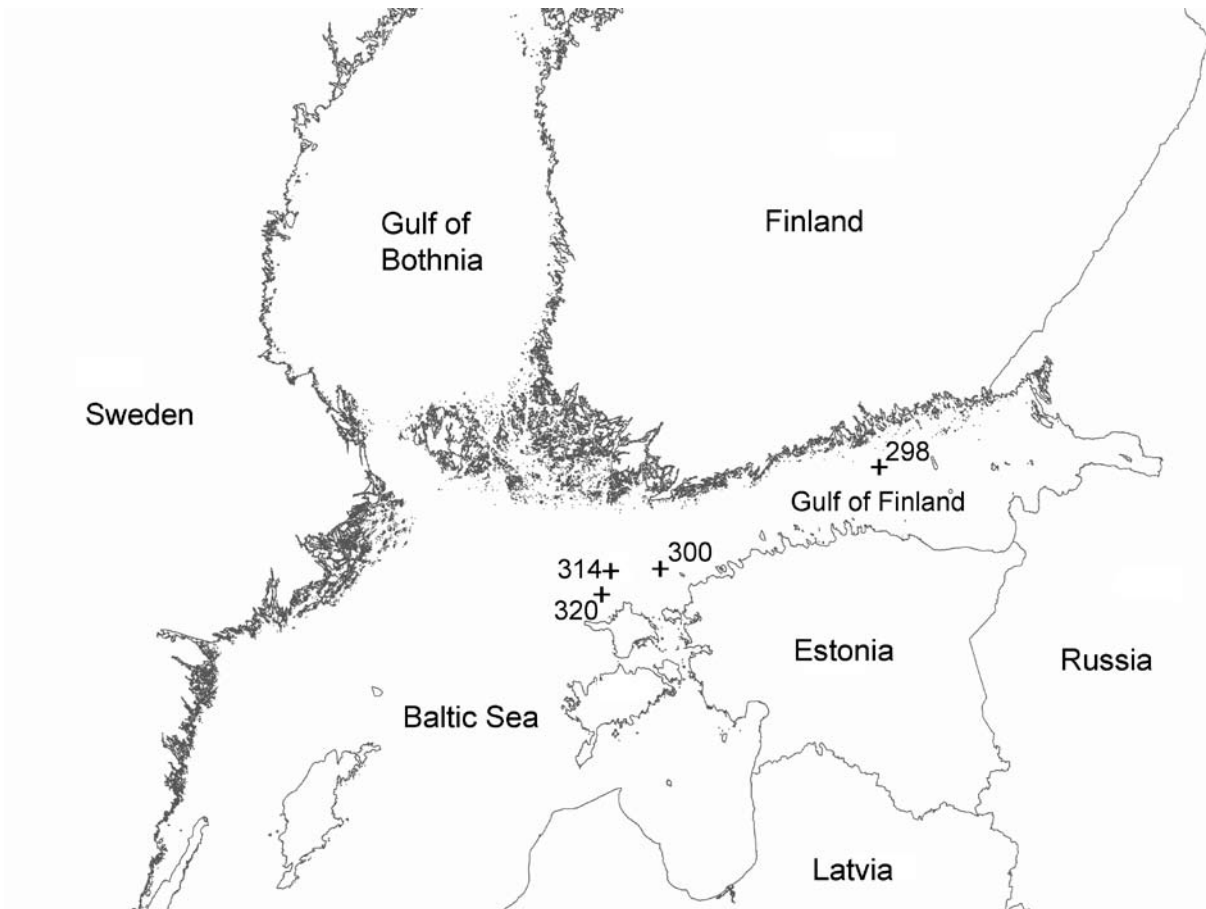
The genus *Synechococcus* is polyphyletic. Several clusters have been identified, based on photosynthetic pigmentation, nitrogen requirements, motility and salinity (Herdman *et al.*, 2001). In marine environments, *Synechococcus* spp. are dominated by members of cluster 5. *Synechococcus* cluster 5 is divided in two sub-clusters, 5.1 and 5.2. Both sub-clusters consist of isolates from the ocean as well as from coastal origin. Members of cluster 5.1 typically have a red colour. They produce PE as their main photosynthetic pigment, have a GC content between 55-62%, and require elevated salt levels for growth. In contrast, members of cluster 5.2 have a green appearance. They produce the pigment PC but lack PE, have a GC content between 63-66%, and are often able to grow without elevated salt requirements (Herdman *et al.*, 2001).

Freshwater picocyanobacteria are often assigned to *Cyanobium*, a genus closely related to *Synechococcus*. *Cyanobium* is only known from freshwater and brackish environments (Crosbie *et al.*, 2003; Ernst *et al.*, 2003). It contains PC as its main photosynthetic pigment and possesses a high GC content (66-71%). *Cyanobium* is composed of clusters that are distinguished by salt-tolerance and GC content (Herdman *et al.*, 2001).

The phylogenetic tree of picocyanobacteria is not always consistent with their pigmentation type. Some strains isolated from marine and freshwater environments produce PE, but are

related to the *Cyanobium* cluster according to sequence information of their 16S rRNA gene, the ribosomal internally transcribed spacer (ITS) region, and the *rpoCI* gene (Crosbie *et al.*, 2003; Ernst *et al.*, 2003; Everroad and Wood, 2006). Conversely, most members of *Synechococcus* cluster 5.1 are rich in PE, but PC-rich isolates were obtained from the Red Sea. Although the genomic GC content of one of these isolates, strain RS9917, (64%) is within the range of *Cyanobium*, it is unknown whether this is also the case for the other strains of that clade (VIII) of cluster 5.1 (Fuller *et al.*, 2003).

Here, we studied natural communities of picocyanobacteria from the Baltic Sea by constructing clone libraries of partial sequences of the 16S rRNA-ITS, *cpeBA* and *cpcBA* operons. The latter two encode for the pigments PE and PC, respectively. Earlier studies suggested that the phylogeny of *cpcBA* of freshwater picocyanobacteria correlated with pigmentation (Neilan *et al.*, 1995; Robertson *et al.*, 2001; Crosbie *et al.*, 2003). Our results demonstrate that a phylogeny based on the operons encoding for phycocyanin and phycoerythrin in picocyanobacteria differs from earlier phylogenies based on the 16S rRNA-ITS operon.



**Figure 3.1.** The sampling stations S298, S300, S314 and S320 along the East-West transect from the Gulf of Finland to the Baltic Sea during the CYANO-cruise in 2004.

## Results

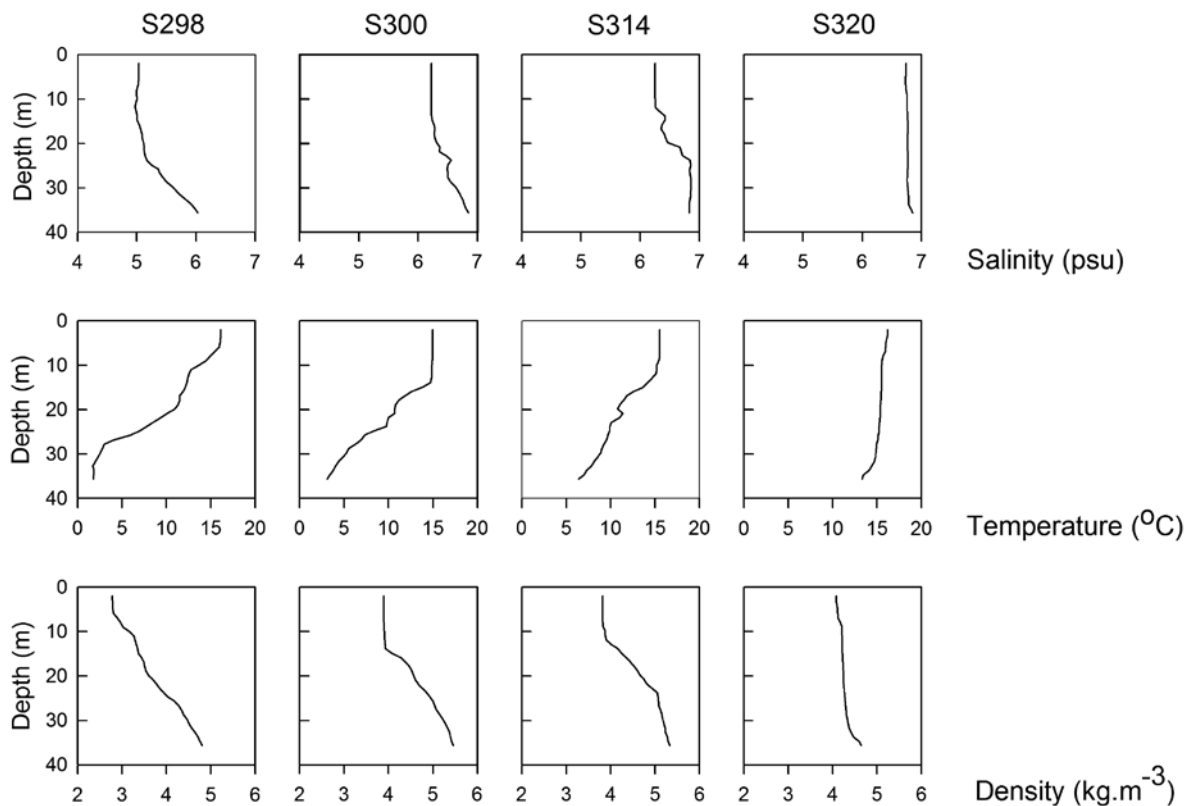
### *Environmental conditions*

#### *Stratification*

In July 2004, water was sampled at 4 stations in the Gulf of Finland and Baltic Sea proper (Fig. 3.1). Station S298 had a triple thermal stratification at 5, 10 and 20 m depth (Fig. 3.2). The density profile of this station revealed that the upper 5 m was well mixed, while density gradually increased with depth below this shallow surface-mixed layer. Station S300 showed a clear surface-mixed layer of ~15 m depth. Station S314 had a slightly shallower surface-mixed layer, with a thermocline and pycnocline at 10-12 m depth. Both stations S300 and S314 had a subtle secondary stratification at ~21 m depth. Station S320 was not stratified, but showed nearly homogeneous vertical profiles of salinity, temperature, and density up to 30 m depth (Fig. 3.2).

#### *Underwater light spectra*

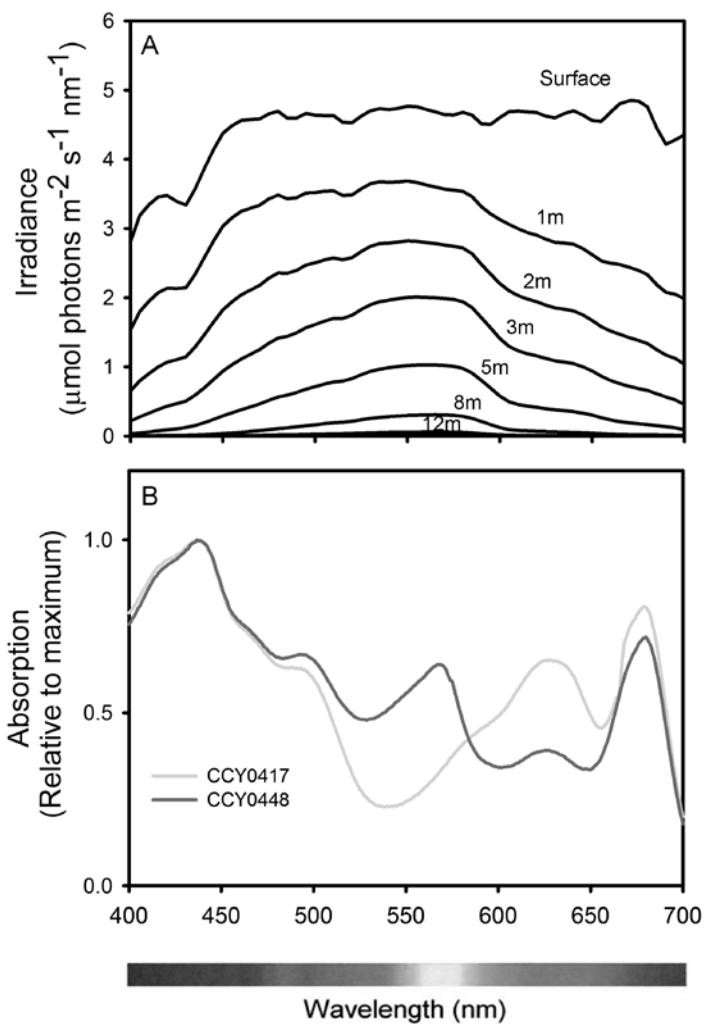
The underwater light spectrum of natural waters largely depends on light attenuation by water itself, by the ‘background turbidity’ caused by dissolved organic matter (known as gilvin in the optics literature) and inanimate suspended particles (tripton, like sediment and



**Figure 3.2.** Vertical profiles of salinity, temperature, and density at the stations S298, S300, S314 and S320.

detritus), and by phytoplankton species present in the water column (Kirk, 1994). Water absorbs strongly in the red part of the spectrum, whereas gilvin and tripton are responsible for rapid attenuation of blue wavelengths. In the Baltic Sea, light absorption in the blue and the red end of the spectrum is of a similar magnitude. At all 4 stations, this yielded an underwater light spectrum that narrowed to green wavelengths with increasing depth (Fig. 3.3A). The light absorption spectra of a red and a green strain of Baltic Sea picocyanobacteria are depicted in Fig. 3.3B as an example to illustrate how they are tuned to the underwater light spectrum. PE-rich strains have an absorption peak at ~560 nm, and hence absorb green light effectively. PC-rich strains have an absorption peak at ~625 nm, and absorb orange-red light effectively. Chlorophyll peaks were also clearly visible in the absorption spectra at 440 nm (Soret band) and 680 nm.

We found euphotic depths of 10.5 m at station S298, 15.3 m at station S300, and 20.3 m at stations S314 and S320, where the euphotic depth is defined as the depth at which the irradiance [PAR( Photosynthetically active radiation), 400-700 nm] equals 1% of the surface irradiance. Hence, the background turbidity of the surface water decreased from the Eastern towards the Western part of the Gulf of Finland.



**Figure 3.3.** Comparison of the underwater light spectrum and the light absorption spectra of PE-rich and PC-rich picocyanobacteria.

**A.** Underwater light spectra measured at station S320 in the (Gulf of Finland (Baltic Sea). The spectrum narrows to the green waveband with increasing depth. Underwater light spectra at the three other stations were similar.

**B.** Absorption spectra of the PC-rich strain CCY0417 and the PE-rich strain CCY0448 isolated from the Gulf of Finland (Baltic Sea). Absorption spectra are scaled to their maximum value.

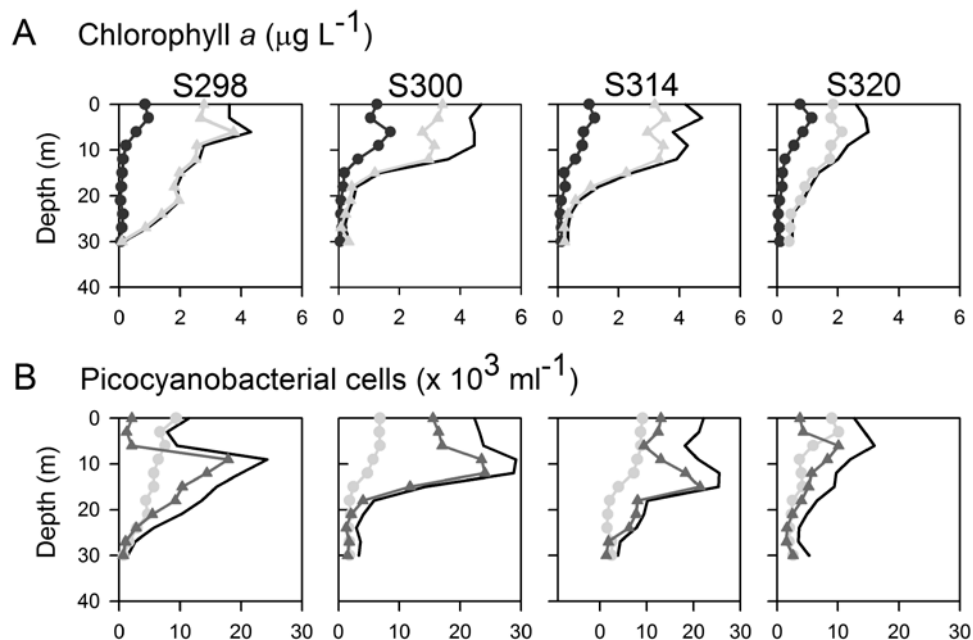
**Table 3.1.** Concentrations of  $PO_4^{3-}$ ,  $NO_3^-$ ,  $NO_2^-$ , and  $NH_4^+$  (in  $\mu\text{mol L}^{-1}$ ) measured at the four sampling stations, both from the surface layer and at 30 m depth.

Stations	$PO_4^{3-}$ ( $\mu\text{M}$ )		$NO_3^-$ ( $\mu\text{M}$ )		$NO_2^-$ ( $\mu\text{M}$ )		$NH_4^+$ ( $\mu\text{M}$ )		N:P	
	0m	30m	0m	30m	0m	30m	0m	30m	0m	30m
S298	0.02	0.28*	0	0.01*	0	0.01*	0.18	0.17*	9	0.68*
S300	0.18	0.85	0	1.11	0	0.29	0.21	0.86	1.17	2.66
S314	0.09	0.32	0	0.32	0	0.08	0.11	1.56	1.22	6.13
S320	0	0.19	0.03	0.09	0	0	0.07	1.3	n.d.	7.32

\*At station S298, deep samples were from 20 m instead of 30m.

### Nutrients

Dissolved inorganic nitrogen and phosphorus were measured in water samples from the surface (0 m) and from 30 m depth (Table 3.1). At all stations, nitrogen and phosphorus concentrations were lower at the surface than at depth. Nitrate and nitrite concentrations at the surface were at or below the detection limit of  $0.01 \mu\text{M}$ . At station S320, the phosphorus concentration at the surface was also below the detection limit. At all stations, the N:P ratios were well below the Redfield ratio of 16 (Table 3.1), indicating that nitrogen was relatively more limiting for phytoplankton growth than phosphorus.



**Figure 3.4.** Vertical profiles of chlorophyll a and picocyanobacteria at stations S298, S300, S314, and S320.

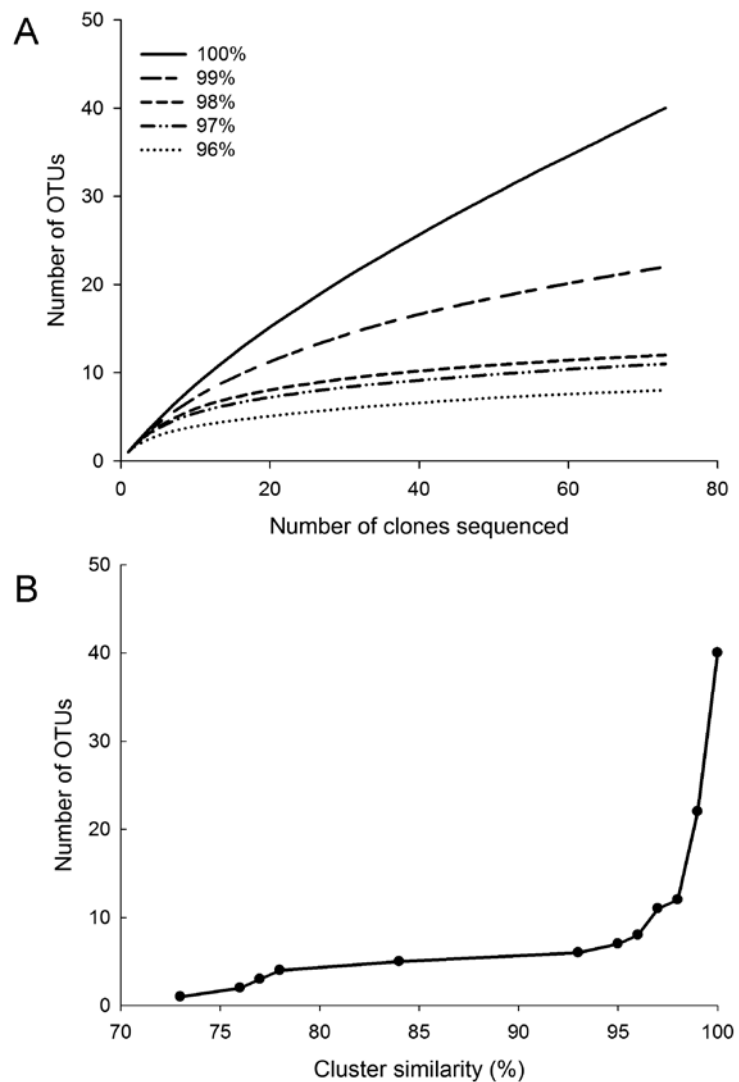
**A.** Concentration of chlorophyll a in the large size fraction (blue dots;  $> 20 \mu\text{m}$ ) and in the small size fraction (green triangles;  $< 20 \mu\text{m}$ ). Total concentration of chlorophyll a is shown in black.

**B.** Concentration of PC-rich picocyanobacteria (green dots) and PE-rich picocyanobacteria (red triangles). In black is shown the total number of picocyanobacterial cells.

### Distribution of picocyanobacteria

Chlorophyll *a* was measured in two size fractions, a small size fraction (< 20  $\mu\text{m}$ ) and a large size fraction (> 20  $\mu\text{m}$ ). Microscopic examination indicated that the small size fraction in the Baltic Sea contained mainly picocyanobacteria (< 2  $\mu\text{m}$ ) and also small filaments of *Pseudanabaena* spp., consistent with earlier studies (Albertano *et al.*, 1997; Stal and Walsby, 2000; Stal *et al.*, 2003). The large size fraction was dominated by the filamentous,  $\text{N}_2$ -fixing cyanobacteria *Nodularia spumigena*, *Anabaena* spp. and *Aphanizomenon flos-aquae*, which were mainly concentrated in the upper 10 m of the water column (Fig. 3.4). Picocyanobacteria were mainly distributed over the upper 15-20 m at stations S300, S314 and S320, and even down to 30 m at station S298. The small size fraction represented 70-80% of the total chlorophyll *a* in the upper 10 m, and even more than 90% of the total chlorophyll *a* below 10 m (Fig. 3.4).

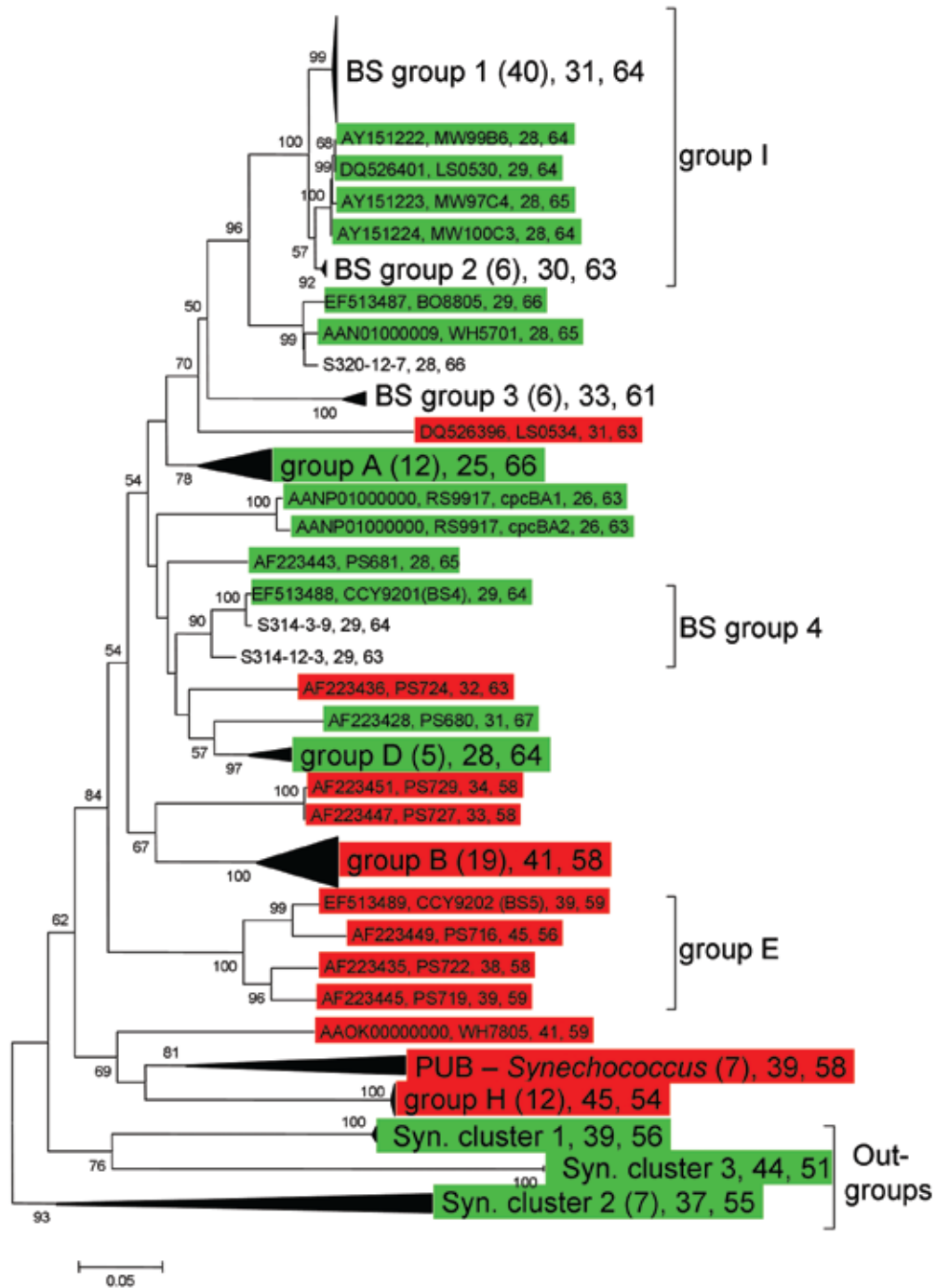
Red and green picocyanobacteria were counted by flow cytometry, on the basis of their size and pigment composition. The depth distributions revealed that red and green picocyanobacteria coexisted throughout the upper 30 m (Fig. 3.4). The cell numbers of the



**Figure 3.5.** Diversity patterns of the Baltic Sea picocyanobacteria using 16S *rRNA*-ITS sequences.

**A.** Rarefaction curves of the number of observed OTUs at 100, 99, 98, 97 and 96 % similarity cut-offs.

**B.** Number of OTUs plotted against different cluster cut-off values in 1.0% increments for sequences grouped into similarity clusters.

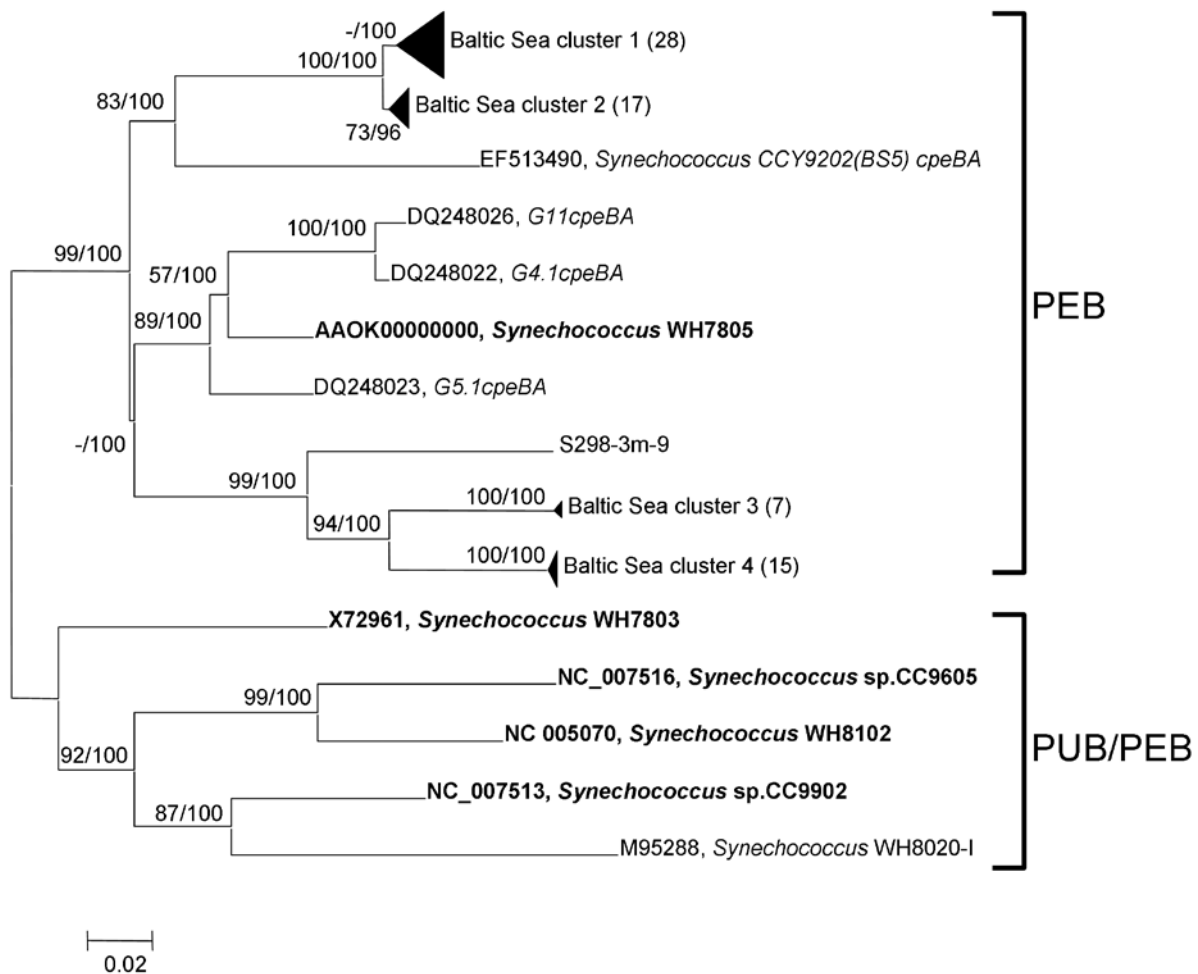


**Figure 3.6.** Neighbor-joining tree of the picocyanobacterial *cpcBA* genes. Clades were condensed for clarity, showing the group designations following Crosbie et al. (2003) (Fig. S2). BS-group designations are assigned to clades formed solely by clone sequences from the Baltic Sea. For condensed groups, the number of *cpcBA* sequences is indicated within brackets. For single sequences, the GenBank accession number and the strain designation are given. For each clade with known isolates, the pigment phenotype is indicated with the colours red (PE-rich) and green (PC-rich). Numbers indicate the mean ENC number and the mean GC content, respectively. The tree was calculated with the software MEGA with the neighbor-joining method using the Kimura- two parameter model of nucleotide substitution with 1000 replicates (Kumar et al., 2004). Bootstrap values (>50%) are shown at the nodes. As out groups were used the *cpcBA* sequences of *Synechococcus* cluster 1 (strains PCC6301, PCC7942 and PCC7943), *Synechococcus* cluster 2 (strains PCC6716, PCC6717, *Synechococcus elongates*, JA-2-3b and JA-3-3b), and *Synechococcus* cluster 3 (PCC7002).

green picocyanobacteria showed a gradual decline with depth, while the red picocyanobacteria formed a subsurface maximum. At stations S298, S300 and S314, the subsurface maximum of the red picocyanobacteria was at the euphotic depth. At station S320, which lacked a clear stratification pattern (Fig. 3.2), the subsurface maximum at ~8 m was less pronounced (Fig. 3.4).

### *The 16S rRNA and ITS region*

The diversity of picocyanobacteria was assessed by sequencing environmental clone libraries containing PCR fragments with a part of the 16S rRNA gene and the internally transcribed



**Figure 3.7.** Unrooted neighbor-joining tree of the picocyanobacterial *cpeBA* genes. Sequences were obtained from the Baltic Sea and from *Synechococcus* strains with sequenced genomes spanning the *cpeBA*-IGS region. Baltic Sea clusters indicate clades formed solely by clone sequences from the Baltic Sea. The number of *cpeBA* clone sequences is indicated within brackets. *Synechococcus* sequences extracted from existing genome sequences or GenBank are shown in bold. Additional *Synechococcus* sequences from strains used in this study are shown in italics. The tree revealed that the *cpeBA* sequences separated into clades containing PEB only and PUB/PEB-producing clades. The Baltic Sea sequences separated into 4 clusters and one single clone (S298-3m-9). Bootstrap values (>50%) based on 1000 replicates are shown at the nodes, using distance analysis (first number) and maximum parsimony analyses (second number). A ‘-’ indicates not significant.

**Table 3.2.** Diversity estimators for the clone libraries of the 16S rRNA-ITS, 16S rRNA, *cpcBA* operon and *cpeBA* operon, with and without intergenic spacers. The number of Operational Taxonomic Units (OTUs) is shown at 100%, 99% and 97% similarity cut-off values. The coverage is expressed as defined by Good (1953). The Chao-1 richness, ACE richness, Shannon diversity index and Simpson diversity index use 99% similarity cut-off values. Numbers within parentheses for the Chao-1 and ACE richness estimators are 95% confidence intervals.

Gene	Number of clones	OTUs (100% / 99% / 97%)	Good's Coverage (%)	Chao-1	S-ACE	Shannon index	Simpson index (1/D)
16S-ITS complete	73	40 / 22 / 11	86.3	37 (26-86)	36 (26-66)	2.64	10.90
16S without ITS	73	19 / 6 / 1	95.9	9 (6-31)	14 (7-79)	0.89	1.85
<i>cpcBA</i> operon	68	24 / 11 / 8	92.65	21 (13-63)	16 (12-37)	1.52	2.76
<i>cpcBA</i> without IGS	68	20 / 10 / 8	94.12	16 (11-48)	13 (11-30)	1.49	2.75
<i>cpeBA</i> operon	68	24 / 11 / 5	91.8	26 (14-79)	28 (14-107)	1.85	5.52
<i>cpeBA</i> without IGS	68	24 / 12 / 6	91.8	27 (15-80)	23 (14-70)	2.01	6.66

spacer between the 16S and 23S rRNA genes (ITS). At all 4 stations, samples were taken at 3 and 12 m depth, where both PC-rich and PE-rich picocyanobacteria were abundant (Fig. 3.4). The samples were size fractionated, to separate the small cyanobacteria (< 20 µm) from the larger phytoplankton. This yielded a total of 8 samples, from which DNA was extracted and PCR amplified using oligonucleotide primers specific for cyanobacteria. We sequenced the last 400 bases of the 16S rRNA gene and the complete ITS of 74 clones, and compared these sequences against existing databases (NCBI, RDP-II) (Table S1, Fig. S1). One clone appeared to be from the filamentous heterocystous cyanobacterium *Anabaena flos-aquae* (99% similarity to the 16S rRNA sequence; AJ630422), and was therefore not further considered.

The vast majority of clones (65 of the 74) exhibited high sequence similarity (96 to 99%) to several closely related *Synechococcus* strains (LM94, BO8807 and *S. rubescens*), which all belong to freshwater group B (Crosbie *et al.*, 2003; Ernst *et al.*, 2003) (Table S1, Fig. S1). This is consistent with earlier studies, which have shown that strains of group B are more than 99% similar at the 16S-rRNA level (Crosbie *et al.*, 2003), and more than 95% similar at the ITS sequence (Ernst *et al.*, 2003). The remaining clones displayed high sequence similarity (96 to 98%) to other freshwater *Synechococcus* strains (Table S1, Fig. S1). One of our clone sequences (TH320-12-6) had a 99% similarity to the 16S rRNA gene of *Synechococcus* strain MH305 (Crosbie *et al.*, 2003). The ITS sequence of this clone was completely disparate from the other clones, except for the tRNA genes. The position of the clone TH320-12-6 in our phylogenetic analysis confirms this by placing the sequence close to the root of the tree with low bootstrap support (Fig. S1). We observed large variations in ITS length and GC content in our clone libraries, consistent with earlier studies (Laloui *et al.*, 2002; Rocap *et al.*, 2002; Ernst *et al.*, 2003; Chen *et al.*, 2006).

Comparison of the clone libraries from 3 m and 12 m depth, using the program WebLibshuff (Singleton *et al.*, 2001), revealed that there was no significant difference between the

libraries obtained from the two sampling depths ( $P > 0.05$ ). We therefore assumed that the libraries from 3 m and 12 m depth have the same composition, and they were lumped in our diversity analysis. The diversity in the clone libraries was analysed using the program DOTUR that calculates several diversity estimators and can be used to create rarefaction curves and similarity plots (Schloss and Handelsman, 2005). Rarefaction was used to determine the diversity structure within the 16S rRNA gene - ITS clone library (Fig. 3.5A, Table 3.2). These results indicate a high degree of microdiversity in our clone library, suggesting that many of the sequences belong to the same or closely related “species”. When the similarity was further reduced, the number of OTUs continued to decrease until all clones merged into a single OTU at 73% similarity (Fig. 3.5B).

Because the ITS region is highly variable, we also tested the diversity within our library by using only the sequences encoding part of the 16S rRNA gene (487 bp). This revealed that 68% of the partial 16S rRNA sequences fall into the 99% clusters (Table 3.2).

Several diversity estimators were calculated, such as the Shannon-Weaver and Simpson diversity indices, Good’s Coverage, and the Chao and ACE richness estimates (Good, 1953; Chao and Lee, 1992; Magurran, 1988). Assuming a 99% similarity criterion, the Chao and ACE richness estimates indicated a species richness of 37 and 36, respectively (Table 3.2).

### *The phycocyanin operon*

We included known *cpcBA* sequences in our alignment for comparison with the 68 clones that we obtained from the Baltic Sea. The lengths of the sequences available in GenBank ranged from 320 bp to almost 500 bp (excluding the intergenic spacer, IGS), complicating phylogenetic analysis of the *cpcBA* genes. We decided to remove sequences shorter than 380 bp (IGS excluded) from our alignment to avoid incorrect topologies (Nei *et al.*, 1998; Tamura *et al.*, 2004). This approach gave a more robust phylogenetic tree of the *cpcBA* gene.

Figure 6 shows the phylogenetic tree that we obtained for the partial *cpcBA* gene sequences. Many of the picocyanobacteria of the Baltic Sea are closely related to the known groups A, B, H, and I (Robertson *et al.*, 2001; Crosbie *et al.*, 2003; Table S2), confirming the results based on the 16S rRNA-ITS operon. The Baltic Sea Group 3 is probably a novel taxon within the picocyanobacteria, since these sequences form a monophyletic group that separates with a long branch and with good bootstrap support from the other sequences. We can not exclude that the other Baltic Sea groups might also represent unique groups although the branch lengths separating these sequences from known sequences are small. Hence, this might as well represent microdiversity between the clusters.

There are also some striking differences between the *cpcBA* phylogeny and the existing 16S rRNA phylogenies (Crosbie *et al.*, 2003; Fuller *et al.*, 2003). First, the *cpcBA* phylogeny separated most picocyanobacteria with a green phenotype from picocyanobacteria with a red phenotype, although there were a few red strains within the green clusters (Fig. 3.6, Fig. S2). Second, in contrast to the 16S rRNA phylogeny, in the *cpcBA* phylogeny green picocyanobacteria isolated from marine environments (e.g., strains RS9917 and WH5701) clustered with green

freshwater picocyanobacteria. Third, the green *Cyanobium* strain CCY9201 (previously known as BS4) and the red *Cyanobium* strain CCY9202 (previously known as BS5), which were nearly identical according to the 16S rRNA-ITS phylogeny (Crosbie *et al.*, 2003; Ernst *et al.*, 2003), were completely separated in the *cpcBA* phylogeny. Fourth, the *cpcBA* phylogeny revealed that phycourobilin (PUB)-producing picocyanobacteria form a distinct cluster within the red picocyanobacteria.

The *cpcBA* phylogeny pointed at a close correlation between pigment phenotype and GC content (Fig. 3.6). PC-rich isolates had GC-contents higher than 60%, while most PE-rich isolates had GC contents less than 60% although there were a few exceptions. The difference in GC content between the *cpcBA* sequences was mainly caused by higher GC content at the third codon position, resulting in synonymous mutations in most of the codons investigated. Likewise, the *cpcBA* phylogeny pointed at a close correlation between pigment phenotype and the effective number of codons (ENC). The ENC number represents a measure for the codon usage bias (Comeron and Aguade, 1998). An ENC number of 20 means that only one codon is used for each amino acid, while an ENC number of 61 indicates that all codons are used equally often and in that case there is no bias in codon usage (Wright, 1990). PC-rich isolates had a low ENC number in the range of 23-32, while almost all PE-rich isolates had a high ENC number ranging from 33 to 45 (Fig. 3.6). Interestingly, PE-rich strains with a GC content exceeding 60% and an ENC-number below 33 clustered with the PC-rich strains.

### *The phycoerythrin operon*

PCR amplification of the *cpeBA* operon encoding the pigment phycoerythrin resulted in 68 clones (for primers see Everroad and Wood, 2006). The number of *cpeBA* sequences available in existing databases such as GenBank was limited to 37 full-length sequences of different cyanobacteria and red algae. BLASTn searches using the nucleotide sequences of all our *cpeBA* clones returned only one of two different top hits, marine *Synechococcus* strains WH7803 (X72961) and WH8102 (BX569694) (Table S3). Our sequences showed only 81% to 90% similarity with these two sequences. BLASTp searches using our *cpeBA* sequences as query were done using the CPE-A and the CPE-B protein coding sequences. Both fragments showed the highest similarity with the CPE-A (range 86 to 93%) and CPE-B (91 to 97%) proteins from the marine *Synechococcus* strain WH7805 (Table S3).

We performed a phylogenetic analysis using our Baltic Sea partial *cpeBA* nucleotide sequences and those recovered from existing databases. Analysis of the phenotypes revealed that all cultured strains within the *cpeBA* phylogeny were PE-rich strains with a GC content between 53 and 63 % and a ENC number ranging from 30 to 45 (Fig. S3). The *cpeBA* phylogeny yielded two major groups (Fig. 3.7). Again these two groups matched the pigmentation of picocyanobacteria. The first group was formed by *cpeBA* genes from freshwater and marine *Synechococcus* strains producing PEB only, while the second group consisted of marine strains producing both PUB and PEB. This topology was consistent with the *cpcBA* phylogeny, where the PUB-producing picocyanobacteria formed a distinct cluster (Fig. 3.6). All *cpeBA* sequences that we obtained from the Baltic Sea were constrained within the PEB group (Fig. 3.7). These

Baltic Sea sequences were separated into two major clades, one clade comprising the clusters 1 and 2, and the other clade formed by clusters 3 and 4. Comparison of the overall similarity at the amino acid level showed that the similarity within each of these two clades is more than 98%, while the similarity between the two clades is only 86.6%.

Calculation of diversity estimators showed that the diversity in the *cpcBA* library and *cpeBA* library is low compared to the 16S rRNA-ITS library (Table 3.2). This might be attributed to inherent differences in variability between these libraries, but also to differences in length between the 16S rRNA-ITS sequences and the *cpcBA* and *cpeBA* sequences. The number of OTUs was rather similar for the *cpcBA* and *cpeBA* operons. According to the Chao-1 and ACE richness estimates and the Shannon and Simpson diversity indices, however, the diversity at the *cpeBA* operon encoding for phycoerythrin was slightly higher than the diversity at the *cpcBA* operon encoding for phycocyanin (Table 3.2).

## Discussion

### *Colourful coexistence of red and green picocyanobacteria*

Our results show that PC-rich and PE-rich picocyanobacteria coexist in the Baltic Sea, where they are approximately equally abundant players in the cyanobacterial community (Fig. 3.4). This confirms earlier results of Stomp *et al.* (2004, 2007). PC-rich picocyanobacteria were slightly more abundant in the upper 5 m of the water column, while PE-rich picocyanobacteria were numerically more dominant at 5-15 m depth. This vertical distribution matches the underwater light spectrum, since green light penetrates more deeply into the Baltic Sea than red light (Fig. 3.3A). Remarkably, the PC-rich and PE-rich picocyanobacteria maintained their vertical distribution even in waters with a nearly homogeneous temperature and density profile (Station S320, Fig. 3.2 and Fig. 3.4). Since picocyanobacteria lack buoyancy regulation, this indicates that the local growth rates of the PE-rich and PC-rich populations at these depths exceeded the rate of vertical mixing by hydrodynamic processes (Huisman *et al.*, 1999).

Sequencing of 209 clones revealed that picocyanobacteria of the Baltic Sea exhibit high levels of microdiversity. Approximately 46% to 54% of the OTUs present in each clone library were constrained at 99% similarity clusters (micro-clusters; Fig. 3.5, Table 3.2). Such high levels of microdiversity have also been detected by many previous studies of marine microbial communities and other natural bacterial populations (Acinas *et al.*, 2004; Lopez-Lopez *et al.*, 2005; Pommier *et al.*, 2007; Rusch *et al.*, 2007). The high microdiversity of *Synechococcus* spp. genes found in our clone libraries may reflect local adaptive radiation of picocyanobacteria which allows them to proliferate under a wide range of different conditions in the Baltic Sea.

### *Phylogeny of red and green picocyanobacteria*

Our results show that a phylogeny based on the *cpcBA* gene (phycocyanin) and *cpeBA* gene (phycoerythrin) differs from a phylogeny based on 16S rRNA gene sequences. This is especially

clear for the *cpcBA* dataset, where clustering of the different phylotypes largely matched the pigment composition of the picocyanobacteria (see also Robertson *et al.*, 2001; Crosbie *et al.*, 2003). This is exemplified by the green CCY9201 (previously known as BS4) and red CCY9202 (previously known as BS5) strains used in the competition experiments of Stomp *et al.* (2004). On the basis of their ITS sequences, these two strains are more than 99% similar (Ernst *et al.*, 2003), whereas their *cpcBA* gene sequences are well separated (Fig. 3.6), where the green strain CCY9201 clusters in the group of PC-rich picocyanobacteria while the red strain CCY9202 clusters in the group of PE-rich picocyanobacteria (Fig. 3.6). The few sequences of red strains that cluster with the *cpcBA* operons of green isolates can be explained by horizontal gene transfer (HGT).

Another example is the placement of the PC-rich marine isolate RS9917. This strain forms a distinct cluster with other PC-rich isolates within the marine picocyanobacteria based on the 16S rRNA gene sequences (Fuller *et al.*, 2003). According to our phylogenetic analysis, the partial *cpcBA* sequences of strain RS9917 clusters with the *cpcBA* sequences of PC-rich freshwater picocyanobacteria. This could have been caused by HGT of the *cpcBA* operon of a freshwater picocyanobacterium. Likewise, clustering of similar pigmentation types is also evident from the placement of PUB/PEB-producing marine *Synechococcus* in both the *cpcBA* and *cpeBA* phylogeny. The marine strain WH7805 produces PEB, but in contrast to other PE-rich marine *Synechococcus* strains it is not capable of producing PUB (Fuller *et al.*, 2003). In the *cpeBA* and *cpcBA* phylogenetic trees, strain WH7805 is clustered separately from the PUB-producing marine *Synechococcus* strains. Only strains that produce PUB might possess the capacity of chromatic adaptation of type IV. We have not retrieved any sequences in our Baltic Sea clone libraries that are related to PUB-producing picocyanobacteria.

Overall, our phylogenetic analyses extend earlier findings of Robertson *et al.* (2001) and Crosbie *et al.* (2003), who showed that the *cpcBA* operon separates PE-rich and PC-rich picocyanobacterial isolates from freshwater lakes. In our analysis, we included picocyanobacteria from brackish waters and marine ecosystems, and studied not only the *cpcBA* operon but also the *cpeBA* operon. This revealed three distinct groups of picocyanobacteria separated in line with their pigmentation, namely PUB/PEB producing strains, PEB producing strains, and PC producing strains. All are members of the monophyletic clade formed by *Synechococcus* and *Cyanobium*.

### *Correlations with GC content and ENC number*

Differences in pigmentation in the *cpcBA* phylogeny correlated with the ENC number and the GC content of the sequences. PC-rich picocyanobacteria had higher GC contents and lower ENC numbers than PE-rich picocyanobacteria (Fig. 3.6). One possible explanation for differences in GC content in PE-rich and PC-rich picocyanobacteria is that it may reflect differences in expression levels of the *cpcBA* gene. In fact, highly expressed genes in *Prochlorococcus* strain MED4 had a higher GC content compared to low expressed genes (Banerjee and Ghosh, 2006). A PE-rich cyanobacterial phycobilisome has one disk of PC proteins while containing multiple disks of PE proteins. A PC-rich phycobilisome usually has

several disks of PC. The higher demand for phycocyanin might require a higher expression level and, hence a higher GC content of the *cpcBA* operon in PC-rich cyanobacteria. Alternatively, it could also be that the genomes of PC-rich picocyanobacteria have a higher GC-content. We tested this hypothesis by analyzing the GC-content of the protein-coding genes of the genome sequences of *Synechococcus* spp. present in GenBank (Table S4). This showed that the overall GC-content of the protein-coding genes of the PC-rich *Synechococcus* strains WH5701 and RS9917 is higher compared to those of the PE-rich picocyanobacteria (Table S4). This would contradict the theory that higher expression levels cause the higher GC-content in the *cpcBA* operons of PC-rich *Synechococcus* spp. It also confirms the placement of RS9917 among the freshwater picocyanobacteria in our phylogenetic analysis and that it is unlikely that this is caused by HGT of phycobiliprotein genes.

Another explanation for the relationship between GC content and pigmentation might come from the environment. Comparative studies suggest that the GC contents of microbial genomes or environmental shotgun libraries vary among habitats of different productivity (Goo *et al.*, 2004; Carbone *et al.*, 2005; Foerstner *et al.*, 2005). For instance, Foerstner *et al.* (2005) observed that the average GC-content of open reading frames (ORFs) from the oligotrophic Sargasso Sea is only 34%, whereas the GC content of ORFs from productive Minnesota soil samples is 61%. These large differences in GC content were not merely an effect of differences in species composition between these two contrasting environments, but remained when the same analysis was focused on phyla present in both environments or on genes present in both environments. Extrapolated to the *cpcBA* phylogeny, this would mean that the high GC sequences of PC-rich picocyanobacteria come from environments with higher levels of nutrients than the sequences of PE-rich picocyanobacteria that have a lower GC content. This explanation is consistent with the global distribution pattern of picocyanobacteria (e.g., Stomp *et al.*, 2007), where PC-rich picocyanobacteria dominate in productive lakes and coastal waters while PE-rich picocyanobacteria dominate in the oligotrophic open ocean.

### Conclusions

We have found high microdiversity among picocyanobacteria of the Baltic Sea, where coexistence of red and green *Synechococcus* strains is widespread. Analysis of the *cpcBA* and *cpeBA* operons revealed a phylogenetic tree in which picocyanobacteria are divided into three different pigment groups: PC-rich, only PEB-producing, and PUB/PEB-producing strains. The PC-rich strains had consistently higher GC contents and lower ENC numbers than the two other pigment groups. These findings differ from the picocyanobacterial phylogeny based on 16S rRNA, which separates marine and freshwater species but not the pigmentation groups. This indicates that picocyanobacterial phylogenies based on the phycocyanin and phycoerythrin genes are not easily compared with the 16S rRNA phylogeny. The topologies can be dissimilar because of different evolutionary histories of the different genes within the same group of organisms.

## Experimental procedures

### *Sample collection*

Water samples from the Baltic Sea were collected from 12 - 19 July 2004 during a research cruise with the Finnish RV *Aranda*. For the work reported here, we sampled 4 stations (stations S298, S300, S314, S320; Fig. 3.1), positioned along an East-West transect from the Gulf of Finland into the Baltic Sea proper (from N 59.1 - 60.0 °N and E 22.2 - 26.2 °E to 59.1 °N 22.2 °E). Samples were taken at 3 m depth intervals from the surface to 30 m depth using a rosette sampler. A Seabird 911 CTD was connected to the rosette sampler, to measure temperature and salinity along these depth profiles. Nutrient concentrations in the water samples were analysed according to standard methods (Grasshoff *et al.*, 1983).

### *Underwater light spectra*

Spectra of the incident light and underwater light spectra were measured with a RAMSES-ACC-VIS spectroradiometer (TriOS, Oldenburg, Germany). Light absorption spectra of isolated strains were measured using a Cary 100 Bio equipped with an integrating sphere DRA-CA-3300, with distilled water as a reference.

### *Chlorophyll analysis*

For chlorophyll *a* analysis, the phytoplankton was divided into two size classes. Total chlorophyll *a* was obtained by filtering 0.5 L on GF/F filters (Whatman, nominal pore size 0.7 µm). Chlorophyll *a* of the large size fraction of phytoplankton was obtained by filtering 1 L on 20 µm nylon mesh (plankton net). Chlorophyll *a* of the small size fraction was calculated as the difference between total chlorophyll *a* and chlorophyll *a* of the large size fraction. This procedure largely discriminates between picoplankton and the larger filamentous cyanobacteria in the Baltic Sea (Stal and Walsby, 2000). Chlorophyll *a* was extracted overnight in the dark at room temperature by 96% ethanol and absorption was measured spectrophotometrically at 665 nm. Chlorophyll concentration was calculated using an absorption coefficient of 72.3 ml mg<sup>-1</sup> cm<sup>-1</sup> (Stal *et al.*, 1999).

### *Counting red and green picocyanobacteria*

The concentrations of red and green picocyanobacteria in the samples were counted by flow cytometry (Jonker *et al.*, 1995; Stomp *et al.*, 2007), using a Coulter Epics Elite ESP flow cytometer (Beckman Coulter Nederland BV, Mijdrecht, Netherlands) equipped with a green laser (525 nm) and a red laser (670 nm). The flow cytometer distinguished between picocyanobacteria and larger phytoplankton by their size (using side scattering). Red and green picocyanobacteria were distinguished based upon their different fluorescence signals. Cells rich in PE emitted orange light (550-620 nm) when excited by the green laser, whereas cells rich in PC emitted far red light (> 670 nm) when excited by the red laser.

### *Extraction of nucleic acids*

From each station 1 L of seawater from each sampling depth was pre-filtered through 20  $\mu\text{m}$  nylon mesh and collected in polycarbonate bottles that were rinsed by 0.5 M NaOH. The pre-filtered seawater was immediately filtered through 0.2  $\mu\text{m}$  Sterivex filtration units (Millipore) using a peristaltic pump. Subsequently, the Sterivex filters were filled with 2 ml lysis buffer (400 mM NaCl, 20 mM EDTA, 50 mM Tris-HCl [pH 9.0], 0.75 M sucrose) (Massana *et al.*, 1997; Moon-van der Staay *et al.*, 2001) and stored at  $-20\text{ }^{\circ}\text{C}$ .

Nucleic acids were extracted as described by Massana *et al.* (1997) with modifications. In brief, lysozyme (final concentration 1 mg ml<sup>-1</sup>) was added to the Sterivex unit and incubated for 45 min at 37  $^{\circ}\text{C}$ . Subsequently, proteinase-K (final concentration 50  $\mu\text{g ml}^{-1}$ ) and sodium dodecyl sulfate (SDS) (1% w/v) were added and incubation was continued overnight at 55  $^{\circ}\text{C}$ . The lysate was recovered from the Sterivex unit by extracting it twice with an equal amount of phenol-chloroform-isoamyl alcohol (25:24:1; pH 8) and once with the same volume of chloroform-isoamyl alcohol (24:1). The extracts were centrifuged (Sigma 4k15 with a swing-out rotor, nr.11156) for 15 min at 1300 rpm and 25  $^{\circ}\text{C}$ . The aqueous phase was transferred to a 15 ml Greiner tube and two volumes of 96% ethanol and 1/10 volume 3 M Na-acetate were added and subsequently incubated for 2 h at  $-70\text{ }^{\circ}\text{C}$  to precipitate the DNA. Subsequently, the DNA was centrifuged for 20 min at 14000 rpm and 4  $^{\circ}\text{C}$ . The pellet was washed with cold 70% ethanol ( $-20\text{ }^{\circ}\text{C}$ ) and centrifuged for 5 min at 14000 rpm and 4  $^{\circ}\text{C}$ . The supernatant was removed by pipetting and the pellet was air dried. The dry pellet was suspended in 100  $\mu\text{l}$  10 mM Tris-HCl (pH 8.5). Because the DNA was not PCR grade after this procedure, it was further purified using the Powersoil DNA extraction kit (MoBio Laboratories) following the manufacturer's recommendations.

### *Primer design*

For amplification of part of the 16S rRNA gene and the internal transcribed spacer between the 16S and 23S rRNA genes, we designed oligonucleotide primers that bind to the 5' region of the 23S rRNA sequences of cyanobacteria (Table 3.3). Cyanobacterial 23S rRNA gene sequences were obtained from GenBank and aligned using the Clustal-W program in Bioedit (Thompson *et al.*, 1994; Hall, 1999). The alignment was imported to Primer Premier software (Premier Biosoft International, version 5.0) and 23S rRNA gene oligonucleotide primers were designed using B1055 as the forward 16S rRNA primer (Singh *et al.*, 1998; Zaballos *et al.*, 2006; Table 3). Primer sequences were checked for their specificity by performing BLASTn searches against the GenBank database.

PCR primers targeting the phycocyanin *cpcBA* operons in a wide range of cyanobacteria were available from the literature (Neilan *et al.*, 1995; Robertson *et al.*, 2001; Crosbie *et al.*, 2003). Recently, genome sequences from a variety of picocyanobacteria became available providing the opportunity to design primers that target specifically the *cpcBA* genes from *Synechococcus*-like cyanobacteria. Using the Integrated Microbial Genomes database (<http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>), *cpcBA* operons were obtained from the following (un-)finished picocyanobacterial genomes: *Synechococcus* PCC6301 (AP008231), PCC7942

**Table 3.3.** Oligonucleotide primers used in this study.

Primer Name	Target gene	Sequence 5' to 3'	Tm (°C)	Reference
B1055	16S rRNA	ATG GCT GTC GTC AGC TCGT	66	Zaballos et al., 2006
Cya23S-58r2	23S rRNA	CGT CCT TCA TCG CCT CTG	58	This study
PITS 1	ITS	TCA GTT GGT AGA GCG CCT GC	56	Ernst et al., 2003
PITS 3	ITS	GTTAGCGGACTCGAACCGC	65	Ernst et al., 2003
SyncpcB-Fw	cpcB	ATGGCTGCTTGCCTGCG	61	This study
SyncpcA-Rev	cpcA	ATCTGGGTGGTGTAGGG	50	This study
B3FW	cpeB	TCA AGG AGA CCT ACA TCG	58	Everroad and Wood, 2006
SynA1R	cpeA	CAG TAG TTG ATC AGR CGC AGG T	64	Everroad and Wood, 2006

(CP000100), CC9311 (CP000435), CC9605 (CP000110), CC9902 (CP000097), RS9917 (AANP01000000), WH5701 (AANO01000000), WH7805 (AAOK01000000), and WH8102 (BX548020) (Markowitz *et al.*, 2006). The *cpcBA* operons M95288 and M95289 from *Synechococcus* strain WH8020 were downloaded from GenBank (Delorimier *et al.*, 1993). The full length *cpcBA* operons were aligned in Bioedit using the ClustalW algorithm. The alignment was imported in Primer Premier 5.0 and used to design primers specifically targeting the *cpcBA* genes from the marine cluster B (*Synechococcus* WH5701) (Table 3.3).

### PCR and clone library construction

DNA obtained from 3 and 12 m depth of stations S298, S300, S314 and S320 were used to amplify the cyanobacterial 16S rRNA-ITS region, the *cpeBA* operon and the *cpcBA* operons using the primers listed in Table 3.3. The PCR reaction mixture was composed of 1  $\mu$ l of template DNA (1 - 20 ng  $\mu$ l<sup>-1</sup>), 2.5  $\mu$ l of 10X PCR buffer (Qiagen), 0.5  $\mu$ l of 10 mM dNTP's mixture (Roche) and 0.62 units of HotStarTaq DNA polymerase (Qiagen). We added 10 pmol of each forward and reverse primer, except for the 16S rRNA-ITS PCR where 5 pmol was used. Sterile MilliQ grade water was added to a final reaction volume of 25  $\mu$ l.

The PCR reactions were run on a GeneAmp System 2700 thermocycler. The program for the 16S-ITS amplification consisted of 15 min hot start at 94 °C; 35 cycles of 1 min at 94 °C; 1 min at 62 °C; and 1 min at 72°C; which was followed by a final elongation step at 72°C for 10 min. For amplification of the *cpeBA* genes the following program was applied: 15 min at 94 °C, 40 cycles of 30 seconds at 94 °C, 30 seconds at 55 °C, and 1.5 min at 72 °C. The final elongation step was 10 min at 72 °C. The same program was used to amplify *cpcBA* except that the elongation step was only 1 min.

PCR-reactions were done in triplicate to decrease variations in amplification (Polz and Cavanaugh, 1998). The PCR products of the triplicate reactions were pooled and cloned. Cloning was done using the TOPO TA cloning kit for sequencing (Invitrogen) following the instructions of the manufacturer. For each sample and PCR product 20 clones were picked

using sterile toothpicks. The cells were transferred to 200  $\mu$ l of sterile LB-Broth and grown overnight. Twenty-five  $\mu$ l of culture was mixed with 25  $\mu$ l of Milli-Q water and heated at 94 °C for 10 minutes. Five  $\mu$ l of the mixture was used for PCR amplification of the insert using the T3 and T7 primers of the vector. Subsequently, 10 positive PCR reactions were chosen per sample and purified using the DNA Clean & Concentrator (Zymo Research). The DNA concentration was measured using a Nanodrop ND1000 (NanoDrop Technologies) spectrophotometer. The PCR product was sequenced using the Big Dye Terminator v1.1 Cycle sequencing kit (Applied Biosystems) according to the manufacturer's instructions. The clones containing *cpeBA* and *cpcBA* fragments were sequenced using the T3 and T7 primers, while the 16S rRNA-ITS clones were sequenced with the primers B1055, Cya23S-58R2, PITS1 and PITS3 (Table 3.3). Sequencing was done with a 3130 Genetic Analyzer (Applied Biosystems). For each clone, the forward and reverse sequences were manually aligned in Bioedit and the sequences were checked against GenBank using BLASTn and BLASTp (Altschul *et al.*, 1990; McGinnis and Madden, 2004). Furthermore, the 16S rRNA clone sequences were compared to the RDP-II database (Cole *et al.*, 2005).

### *Diversity calculations and phylogenetic analysis*

For the diversity calculations, the clone sequences of the different sampling stations were grouped together. The program DOTUR was used for calculating rarefaction, library coverage, Shannon-Weaver diversity index ( $H'$ ), Simpson index ( $D$ ), Chao-1 non-parametric richness estimator and the ACE coverage-based richness estimator (Schloss and Handelsman, 2005). Calculations were performed on a Jukes-Cantor corrected distance matrix created with the DNADIST program from the PHYLIP Package (Felsenstein, 1989).

Sequences previously identified to be closely related by BLASTn comparison were imported from GenBank into Bioedit and aligned against the clone sequences using ClustalW. Alignments of the 16S rRNA-ITS sequences were done manually in Bioedit by reference of the ITS alignment of the predicted secondary structure models proposed in several papers describing the cyanobacterial ITS sequences (Iteman *et al.*, 2000; Laloui *et al.*, 2002; Rocap *et al.*, 2002; Taton *et al.*, 2003). Sequence comparison and phylogenetic analyses were performed using the software MEGA3.1 (Kumar *et al.*, 2004). For the 16S rRNA-ITS region the sequences were compared using the neighbor-joining algorithm with Jukes-Cantor correction and 1000 bootstraps. The coding regions of the *cpeBA* and *cpcBA* operon were both used in phylogenetic analyses. Both data sets were separately analysed using the following approach. Phylogenetic analyses were done with the neighbor-joining method as well as with maximum parsimony. Neighbor-joining was performed with the Kimura-2-parameter model for nucleotide evolution with 1000 bootstraps. Maximum parsimony was used with the close-neighbor-interchange search algorithm with random tree addition using 100 bootstraps. Codon usage in the *cpcBA* and *cpeBA* coding regions was analysed using DnaSP version 4.0 (Rozas *et al.*, 2003).

### *Nucleotide sequence accession numbers*

The sequence data reported in this paper have been submitted to the GenBank database under accession numbers: 16S rRNA- ITS clones (EF513279 – EF513350); *cpcBA* clones (EF513351 – EF513418); *cpeBA* clones (EF513418 – EF513486); BO8805 *cpcBA* (EF513487); CCY9201 *cpcBA* (EF513488); CCY9202 *cpcBA* (EF513489); CCY9202 *cpeBA* (EF513490); *Anabaena*-like 16S-ITS clone TH298-12-6 (EF530539).

### **Acknowledgements**

We thank M. Laamanen for the opportunity to join cruise CYANO-04, and the crew of the research vessel Aranda for help during sampling. We also thank A. Wijnholds-Vreman for carrying out the flow cytometry analyses. We thank C. Everroad and A.M. Wood for sharing with us their *cpeBA* primer sequences before publication. We gratefully acknowledge the comments of two anonymous referees. M.S. and J.H. were supported by the Earth and Life Sciences Foundation (ALW), which is subsidised by the Netherlands Organization for Scientific Research (NWO). T.H. and L.J.S. acknowledge support from the European Commission through the project MIRACLE (EVK3-CT-2002–00087). This is NIOO-KNAW publication nr: 4153.

## SUPPLEMENTARY MATERIAL CHAPTER 3

The following supplementary material is available for this article:

**Figure S1.** Full length neighbor-joining tree of the picocyanobacterial ITS sequences.

**Figure S2.** Neighbor-joining tree of the partial sequences of picocyanobacterial *cpcBA* genes, showing the positions of all *cpcBA* clones obtained from the Baltic Sea.

**Figure S3.** Neighbor-joining tree of the partial sequences of picocyanobacterial *cpeBA* genes, showing the positions of all *cpeBA* clones obtained from the Baltic Sea.

**Table S1.** Comparison of 16S rRNA- ITS clones with GenBank and RDPII databases.

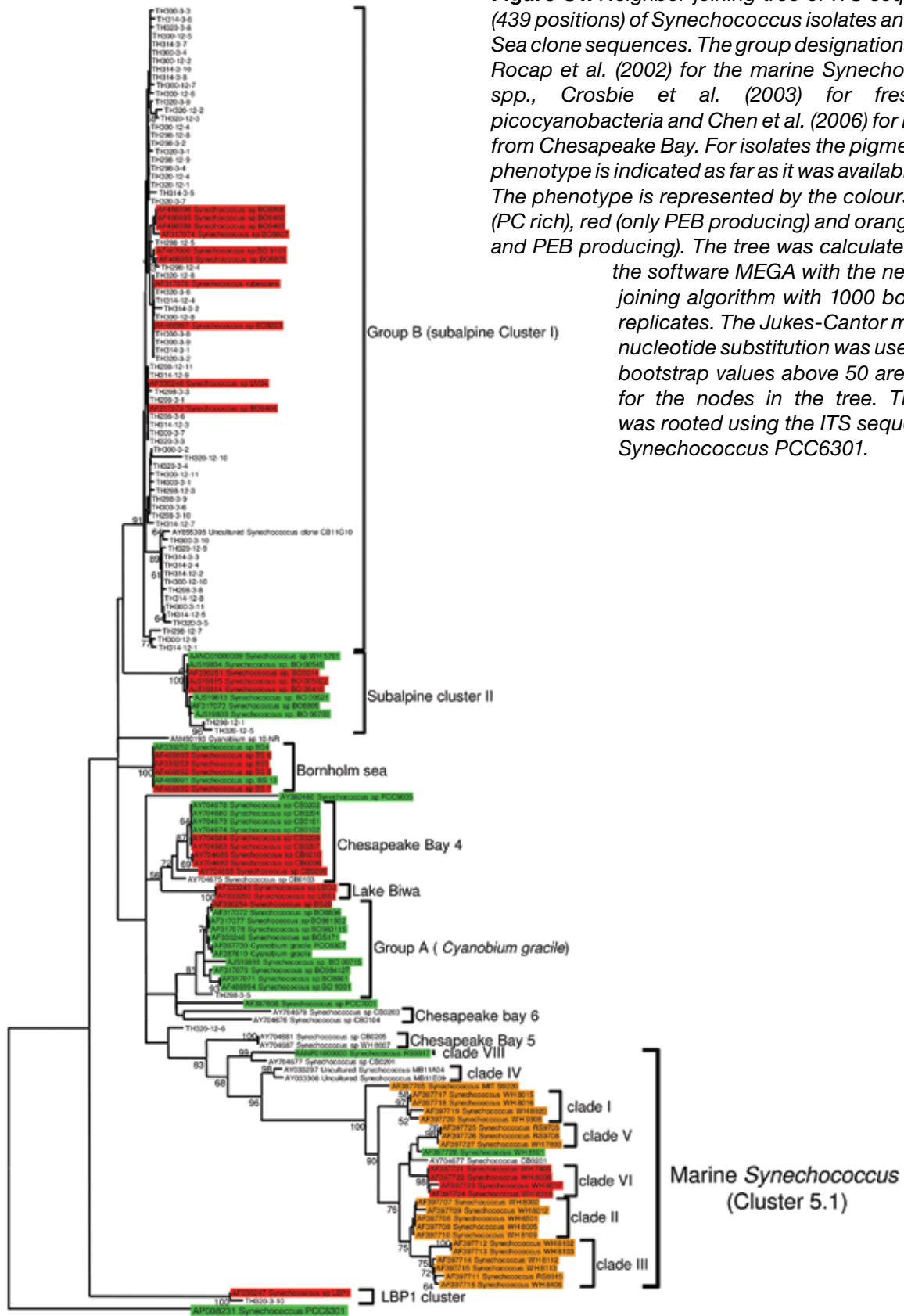
**Table S2.** Comparison of *cpcBA* clones with GenBank databases.

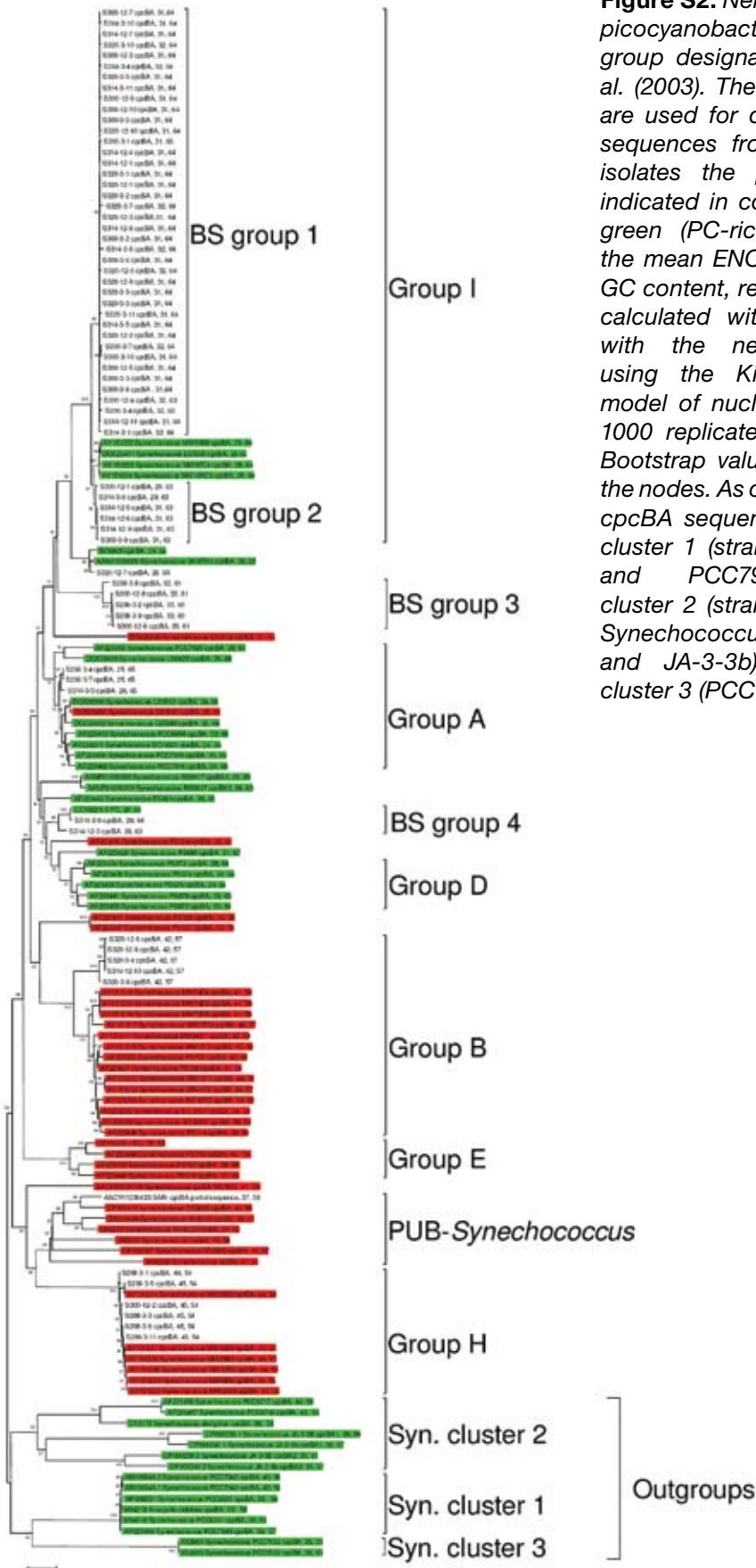
**Table S3.** Comparison of *cpeBA* clones with GenBank databases.

**Table S4a, Table S4b.** Comparison of GC-content from cyanobacterial genomes and the protein coding genes found in these genomes.

## References

- Chen, F., Wang, K., Kan, J.J., Suzuki, M.T., and Wommack, K.E. (2006) Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* **72**: 2239-2243.
- Crosbie, N.D., Pöckl, M., and Weisse, T. (2003) Dispersal and phylogenetic diversity of nonmarine picocyanobacteria, inferred from 16S rRNA gene and *cpcBA*-intergenic spacer sequence analyses. *Appl Environ Microbiol* **69**: 5716-5721.
- Kumar, S., Tamura, K., and Nei, M. (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**: 150-163.
- Rocap, G., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68**: 1180-1191.



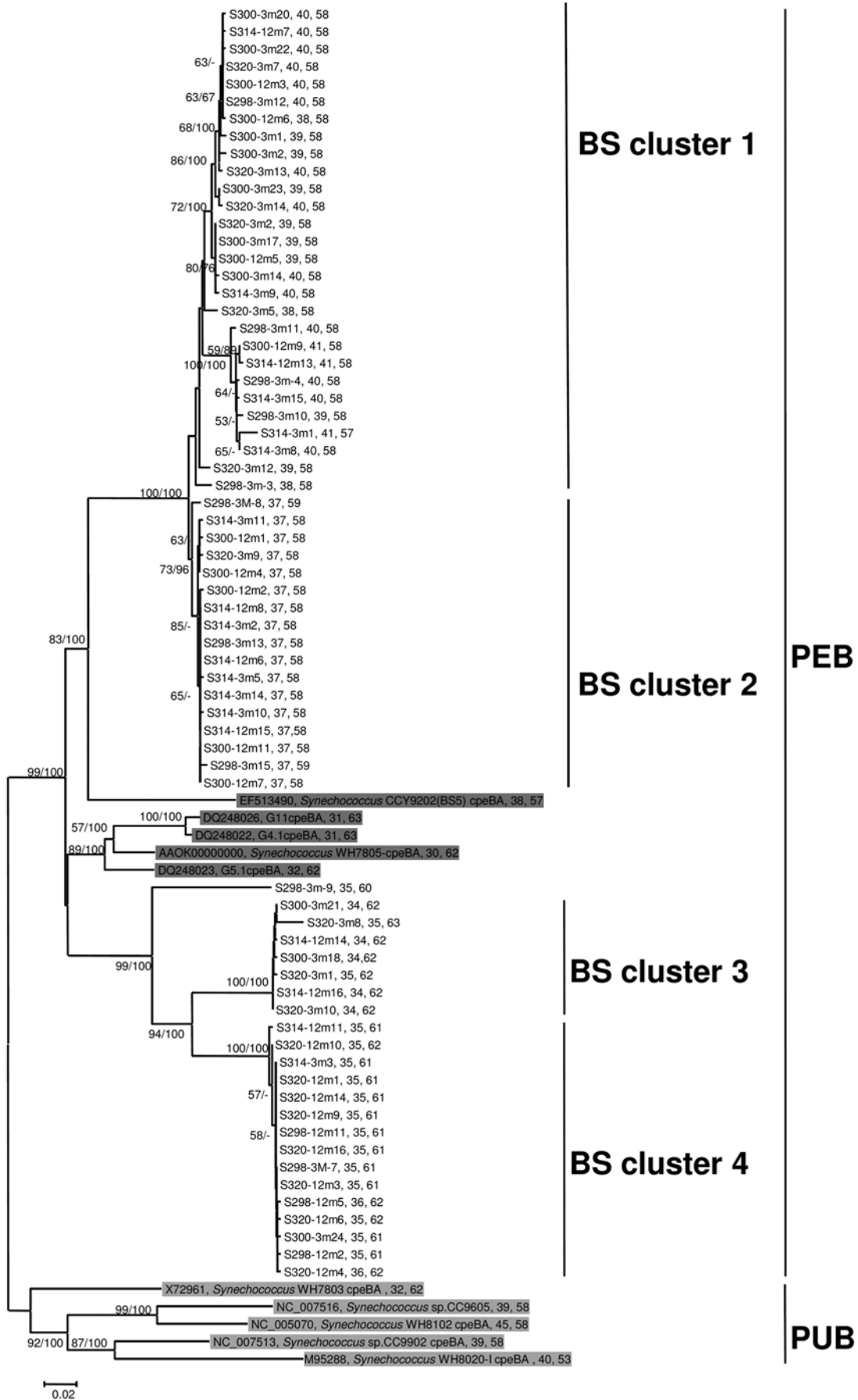


**Figure S2:** Neighbor-joining tree of the picocyanobacterial *cpcBA* genes. The group designations follow Crosbie et al. (2003). The BS-group designations are used for clades formed solely by sequences from the Baltic Sea. For isolates the pigment phenotype is indicated in colours red (PE-rich) and green (PC-rich). Numbers indicated the mean ENC number and the mean GC content, respectively. The tree was calculated with the software MEGA with the neighbor-joining method using the Kimura- two parameter model of nucleotide substitution with 1000 replicates (Kumar et al., 2004). Bootstrap values >50% are shown at the nodes. As out groups were used the *cpcBA* sequences of *Synechococcus* cluster 1 (strains PCC6301, PCC7942 and PCC7943) *Synechococcus* cluster 2 (strains PCC6716, PCC6717, *Synechococcus elongatus*, JA-2-3b and JA-3-3b) and *Synechococcus* cluster 3 (PCC7002).

Table S1: comparison of 16S rRNA- ITS-1 clones with Genbank and RDP-II databases.

Sequence ID	Sampling Location	Depth (m)	GenBank Accession Number	Length (bp)	Closest relative to GenBank (BlastN)	GenBank Acc. Number	Similarity Score (%)	Best hit RDP-II query (only 16S rRNA)	GenBank Acc. Number	RDP-II Score	Phylogenetic clade*
T-298-3-1	60.04° N 26.21° E	3	F513279	1412	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW77D1	AV151246	1.000	B
T-298-3-2	60.04° N 26.21° E	3	F513280	1412	Synechococcus sp. BO8807	AF317074	96%	Synechococcus sp. MW77D1	AV151246	0.985	B
T-298-3-3	60.04° N 26.21° E	3	F513281	1412	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW15-2	AV151240	0.983	B
T-298-3-4	60.04° N 26.21° E	3	F513282	1412	Synechococcus sp. LM94	AF330248	97%	Synechococcus sp. MW72C6	AV151246	1.000	A
T-298-3-5	60.04° N 26.21° E	3	F513283	1437	Synechococcus sp. BO983115	AF317074	96%	Synechococcus sp. EG5171	AV151246	1.000	B
T-298-3-6	60.04° N 26.21° E	3	F513284	1482	Uncultured Synechococcus sp. clone CB11G10	AF317074	96%	Synechococcus sp. MW77D1	AV151246	0.985	B
T-298-3-7	60.04° N 26.21° E	3	F513285	1392	Uncultured Synechococcus sp. clone CB11G10	AF317074	96%	Synechococcus sp. MW72C6	AV151240	0.998	B
T-298-3-8	60.04° N 26.21° E	3	F513286	1412	Synechococcus sp. BO8807	AF317074	96%	Synechococcus sp. MW72C6	AV151240	0.998	B
T-298-3-9	60.04° N 26.21° E	3	F513287	1357	Synechococcus sp. BO8807	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.996	B
T-298-12-1	60.04° N 26.21° E	12	F513288	1412	Synechococcus sp. BO8807	AF330248	97%	Synechococcus sp. MW77D1	AV151246	1.000	B
T-298-12-2	60.04° N 26.21° E	12	F513289	1412	Synechococcus sp. BO8807	AF317074	96%	Synechococcus sp. MW77D1	AV151246	0.992	B
T-298-12-3	60.04° N 26.21° E	12	F513290	1412	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW15-2	AV151240	0.972	B
T-298-12-4	60.04° N 26.21° E	12	F513291	1412	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW15-2	AV151240	0.972	B
T-298-12-5	60.04° N 26.21° E	12	F513292	1412	Anabaena flos-aquae 1LJ3064	AJ630422	99%	Aphanizomenon flos-aquae var. klebahnii. 218	AV151246	0.987	B
T-298-12-6	60.04° N 26.21° E	12	F513293	1439	Synechococcus sp. BO8807	AF317074	96%	Synechococcus sp. MW77D1	AV151246	0.989	B
T-298-12-7	60.04° N 26.21° E	12	F513294	1412	Synechococcus sp. BO8807	AF317074	96%	Synechococcus sp. MW77D1	AV151246	0.955	B
T-298-12-8	60.04° N 26.21° E	12	F513295	1438	Synechococcus tubescens	AF330248	97%	Synechococcus sp. MW77D1	AV151246	1.000	B
T-298-12-9	60.04° N 26.21° E	12	F513296	1412	Synechococcus sp. BO8807	AF317074	96%	Synechococcus sp. MW77D1	AV151246	0.991	B
T-298-12-10	60.04° N 26.21° E	12	F513297	1412	Synechococcus tubescens	AF330248	96%	Synechococcus sp. MW77D1	AV151246	0.985	B
T-298-12-11	60.04° N 26.21° E	12	F513298	1412	Synechococcus tubescens	AF317074	96%	Synechococcus sp. MW15-2	AV151240	0.991	B
T-300-3-1	59.33° N 23.10° E	3	F513300	1366	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW15-2	AV151240	1.000	B
T-300-3-2	59.33° N 23.10° E	3	F513301	1365	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.985	B
T-300-3-3	59.33° N 23.10° E	3	F513302	1438	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.960	B
T-300-3-4	59.33° N 23.10° E	3	F513303	1412	Synechococcus tubescens	AF330248	97%	Synechococcus sp. MW72C6	AV151240	0.977	B
T-300-3-5	59.33° N 23.10° E	3	F513304	1412	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW77D1	AV151246	1.000	B
T-300-3-6	59.33° N 23.10° E	3	F513305	1412	Synechococcus tubescens	AF317076	97%	Synechococcus sp. MW77D1	AV151246	0.987	B
T-300-3-7	59.33° N 23.10° E	3	F513306	1412	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW15-2	AV151240	0.991	B
T-300-3-8	59.33° N 23.10° E	3	F513307	1440	Synechococcus sp. BO8807	AF317074	97%	Synechococcus sp. MW76B2	AV151245	0.970	B
T-300-3-9	59.33° N 23.10° E	3	F513308	1366	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.985	B
T-300-3-10	59.33° N 23.10° E	3	F513309	1412	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.970	B
T-300-3-11	59.33° N 23.10° E	3	F513310	1412	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.960	B
T-300-12-1	59.33° N 23.10° E	12	F513311	1412	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW77D1	AV151246	1.000	B
T-300-12-2	59.33° N 23.10° E	12	F513312	1412	Synechococcus tubescens	AF317076	97%	Synechococcus sp. MW77D1	AV151246	0.987	B
T-300-12-3	59.33° N 23.10° E	12	F513313	1412	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW15-2	AV151240	0.991	B
T-300-12-4	59.33° N 23.10° E	12	F513314	1440	Synechococcus sp. BO8807	AF317074	97%	Synechococcus sp. MW76B2	AV151245	0.970	B
T-300-12-5	59.33° N 23.10° E	12	F513315	1366	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.985	B
T-300-12-6	59.33° N 23.10° E	12	F513316	1412	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW72C6	AV151240	1.000	B
T-300-12-7	59.33° N 23.10° E	12	F513317	1412	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW15-2	AV151240	0.977	B
T-300-12-8	59.33° N 23.10° E	12	F513318	1366	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.977	B
T-300-12-9	59.33° N 23.10° E	12	F513319	1366	Synechococcus sp. BO8807	AF330248	96%	Synechococcus sp. BO 9807	AV151245	0.991	B
T-300-12-10	59.33° N 23.10° E	12	F513320	1438	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.962	B
T-300-12-11	59.33° N 23.10° E	12	F513321	1438	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.957	B
T-300-12-12	59.33° N 23.10° E	12	F513322	1412	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW77D1	AV151246	0.970	B
T-300-12-13	59.33° N 23.10° E	12	F513323	1412	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW77D1	AV151246	0.970	B
T-300-12-14	59.33° N 23.10° E	12	F513324	1440	Synechococcus tubescens	AF317074	97%	Synechococcus sp. MW76B2	AV151245	0.960	B
T-300-12-15	59.33° N 23.10° E	12	F513325	1366	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.972	B
T-300-12-16	59.33° N 23.10° E	12	F513326	1412	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW77D1	AV151246	1.000	B
T-300-12-17	59.33° N 23.10° E	12	F513327	1412	Synechococcus tubescens	AF317074	96%	Synechococcus sp. MW15-2	AV151240	0.985	B
T-300-12-18	59.33° N 23.10° E	12	F513328	1391	Synechococcus tubescens	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.970	B
T-300-12-19	59.33° N 23.10° E	12	F513329	1438	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.977	B
T-300-12-20	59.33° N 23.10° E	12	F513330	1364	Synechococcus sp. BO8807	AF330248	96%	Synechococcus sp. MW76B2	AV151245	0.985	B
T-300-12-21	59.33° N 23.10° E	12	F513331	1412	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.985	B
T-300-12-22	59.33° N 23.10° E	12	F513332	1458	Synechococcus sp. LM94	AF330248	97%	Synechococcus sp. MW77D1	AV151246	1.000	B
T-300-12-23	59.33° N 23.10° E	12	F513333	1458	Synechococcus tubescens	AF317076	96%	Synechococcus sp. MW15-2	AV151240	0.985	B
T-300-12-24	59.33° N 23.10° E	12	F513334	1412	Synechococcus tubescens	AF330248	96%	Synechococcus sp. MW77D1	AV151246	1.000	B
T-300-12-25	59.33° N 23.10° E	12	F513335	1385	Synechococcus sp. LM94	AF317074	99%	Synechococcus sp. MW72C6	AV151240	1.000	B
T-300-12-26	59.33° N 23.10° E	12	F513336	1412	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.972	B
T-300-12-27	59.33° N 23.10° E	12	F513337	1385	Synechococcus sp. BO8807	AF317074	99%	Synechococcus sp. MW76B2	AV151245	0.972	B
T-300-12-28	59.33° N 23.10° E	12	F513338	1412	Synechococcus tubescens	AF317074	99%	Synechococcus sp. MW15-2	AV151240	0.987	B
T-300-12-29	59.33° N 23.10° E	12	F513339	1439	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.987	B
T-300-12-30	59.33° N 23.10° E	12	F513340	1439	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.962	B
T-300-12-31	59.33° N 23.10° E	12	F513341	1306	Synechococcus sp. LBPT1	AF330248	99%	Synechococcus sp. MW25B3	AV151240	1.000	B
T-300-12-32	59.33° N 23.10° E	12	F513342	1438	Synechococcus sp. LBPT1	AF317076	97%	Synechococcus sp. MW72C6	AV151240	0.974	B
T-300-12-33	59.33° N 23.10° E	12	F513343	1438	Synechococcus tubescens	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.985	B
T-300-12-34	59.33° N 23.10° E	12	F513344	1438	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.964	B
T-300-12-35	59.33° N 23.10° E	12	F513345	1437	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.964	B
T-300-12-36	59.33° N 23.10° E	12	F513346	1362	Synechococcus sp. LM94	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.966	B
T-300-12-37	59.33° N 23.10° E	12	F513347	1362	Synechococcus sp. BO8807	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.966	B
T-300-12-38	59.33° N 23.10° E	12	F513348	1362	Synechococcus sp. BO8807	AF330248	96%	Synechococcus sp. MW72C6	AV151240	0.966	B
T-300-12-39	59.33° N 23.10° E	12	F513349	1412	Synechococcus tubescens	AF224198	99%	Synechococcus sp. MH305	AV151240	0.991	B
T-300-12-40	59.33° N 23.10° E	12	F513350	1391	Uncultured Synechococcus sp. clone CB11G10	AF317076	96%	Synechococcus sp. MW76B2	AV151245	1.000	B
T-300-12-41	59.33° N 23.10° E	12	F513351	1412	Synechococcus sp. BO8807	AF317074	96%	Synechococcus sp. MW72C6	AV151240	1.000	B

\* Phylogenetic clade designations follow Crosbie et al., 2003 and Ernst et al., 2003



**Figure S3** (previous page): Unrooted neighbor-joining tree of *cpeBA* sequences obtained from the Baltic Sea and sequences from *Synechococcus* strains spanning the *cpeBA*-IGS region. Isolates producing only PEB are shown in red. The PUB/PEB producing isolates are shown in orange. Numbers indicated the mean ENC number and the mean GC content, respectively. The tree revealed that the *cpeBA* sequences separated into clades containing PEB only and PUB/PEB producing clades. The Baltic Sea sequences separated into 4 clusters and one single clone (S298-3m-9). Significant (<50%) percentage values of 1000 bootstraps are given as numbers at branching points for distance / maximum parsimony analyses (in order NJ-distance / MP) ('-' indicates not significant).



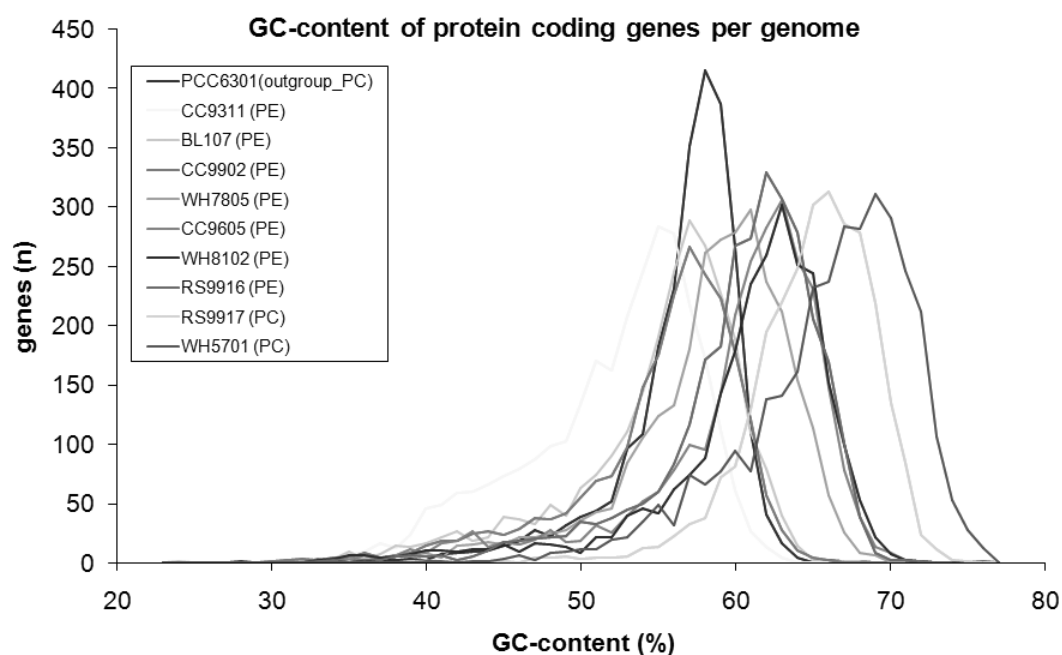


**Table S4a:** GC content of picocyanobacterial genomes

Data obtained from the IMG database: <http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>.  
Genomes organized according to GC-content.

Genome Name	Mbp	Total Genes	Protein coding genes	genomic GC content %	pheno-type	PUB	Site of isolation
<i>Synechococcus elongatus</i> PCC 6301 (outgroup)	2.7	2578	2527	55	PC-rich	n.a.	Waller Creek, Austin, Texas, U.S.A.
<i>Synechococcus</i> sp. CC9311	2.6	2942	2892	52	PE-rich	Yes	California current
<i>Synechococcus</i> sp. BL107	2.3	2553	2507	54	PE-rich	Yes*	Blanes Bay, Mediterranean Sea
<i>Synechococcus</i> sp. CC9902	2.2	2358	2307	54	PE-rich	Yes	California current
<i>Synechococcus</i> sp. WH 7805	2.6	2934	2883	58	PE-rich	No	Sargasso Sea, 33°45'N, 67°30'W
<i>Synechococcus</i> sp. CC9605	2.5	2753	2702	59	PE-rich	Yes	California current
<i>Synechococcus</i> sp. WH 8102	2.4	2580	2528	59	PE-rich	Yes	Tropical Atlantic, 22°30'N, 65°36'W
<i>Synechococcus</i> sp. RS9916	2.7	3009	2961	60	PE-rich	Yes	Gulf of Aqaba, 29°28'N, 34°55'E
<i>Synechococcus</i> sp. RS9917	2.6	2820	2770	64	PC-rich	n.a.	Gulf of Aqaba, 29°28'N, 34°55'E
<i>Synechococcus</i> sp. WH 5701	3.0	3401	3346	65	PC-rich	n.a.	Long Island Sound

\* D.Scanlan Personal Communication



**Table S4b:** GC content of protein coding genes of PE-rich and PC-rich *Synechococcus* genomes  
Data used to produce the above plots.

GC content per gene (%)	PCC6301 (outgroup_PC)	CC9311 (PE)	BL107 (PE)	CC9902 (PE)	WH7805 (PE)	CC9605 (PE)	WH8102 (PE)	RS9916 (PE)	RS9917 (PC)	WH5701 (PC)
23	0	0	0	0	0	0	0	0	0	0
24	0	0	1	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	1	0	0	0
29	0	0	0	0	0	0	0	0	0	0
30	0	0	0	1	0	0	0	0	0	0
31	0	2	0	3	0	0	0	0	0	0
32	0	3	1	4	0	0	0	0	0	0
33	0	1	3	3	2	4	2	1	0	0
34	1	1	4	5	1	2	4	2	0	0
35	1	7	10	5	4	3	7	1	1	0
36	4	10	4	9	2	7	7	2	0	1
37	2	17	8	5	3	4	5	2	0	0
38	3	13	15	7	7	7	6	6	0	0
39	4	24	14	9	6	12	8	7	1	0
40	3	46	17	15	9	9	11	6	2	1
41	8	49	22	19	5	17	11	7	1	1
42	9	59	27	19	15	19	10	2	0	0
43	10	60	19	25	16	27	12	6	1	1
44	11	66	22	27	18	14	12	11	0	1
45	15	73	39	24	16	18	16	18	1	4
46	19	79	37	29	22	16	10	19	0	7
47	28	88	33	38	22	20	17	21	5	3
48	23	99	49	37	25	28	16	22	5	9
49	32	103	40	42	28	18	14	23	6	10
50	39	134	63	55	36	19	9	35	4	12
51	44	171	74	69	43	32	22	33	5	12
52	52	162	91	73	46	26	22	38	5	20
53	97	208	111	98	85	40	40	45	6	22
54	109	245	144	148	106	52	46	50	13	35
55	183	284	191	176	124	60	42	60	14	49
56	231	278	245	226	133	78	62	83	23	32
57	352	218	289	267	181	100	74	117	33	74
58	416	173	268	243	262	96	89	172	38	66
59	387	112	227	223	273	143	143	183	72	77
60	266	61	189	178	279	209	179	268	81	95
61	111	27	110	121	298	255	235	274	130	77
62	41	14	79	57	237	283	260	330	195	138
63	17	5	40	28	212	306	302	307	220	141
64	5	0	15	11	154	256	252	279	249	161
65	0	0	2	5	114	231	244	206	302	232
66	2	0	2	3	57	148	152	170	313	237
67	1	0	1	0	26	78	100	101	286	284
68	1	0	1	0	9	39	53	39	279	282
69	0	0	0	0	7	14	22	9	219	311
70	0	0	0	0	0	9	8	2	135	291
71	0	0	0	0	0	3	2	2	85	246
72	0	0	0	0	0	0	1	1	24	212
73	0	0	0	0	0	0	0	0	12	107
74	0	0	0	0	0	0	0	0	3	53
75	0	0	0	0	0	0	0	0	1	28
76	0	0	0	0	0	0	0	1	0	13
77	0	0	0	0	0	0	0	0	0	0

