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

Haverkamp, T.H.A.

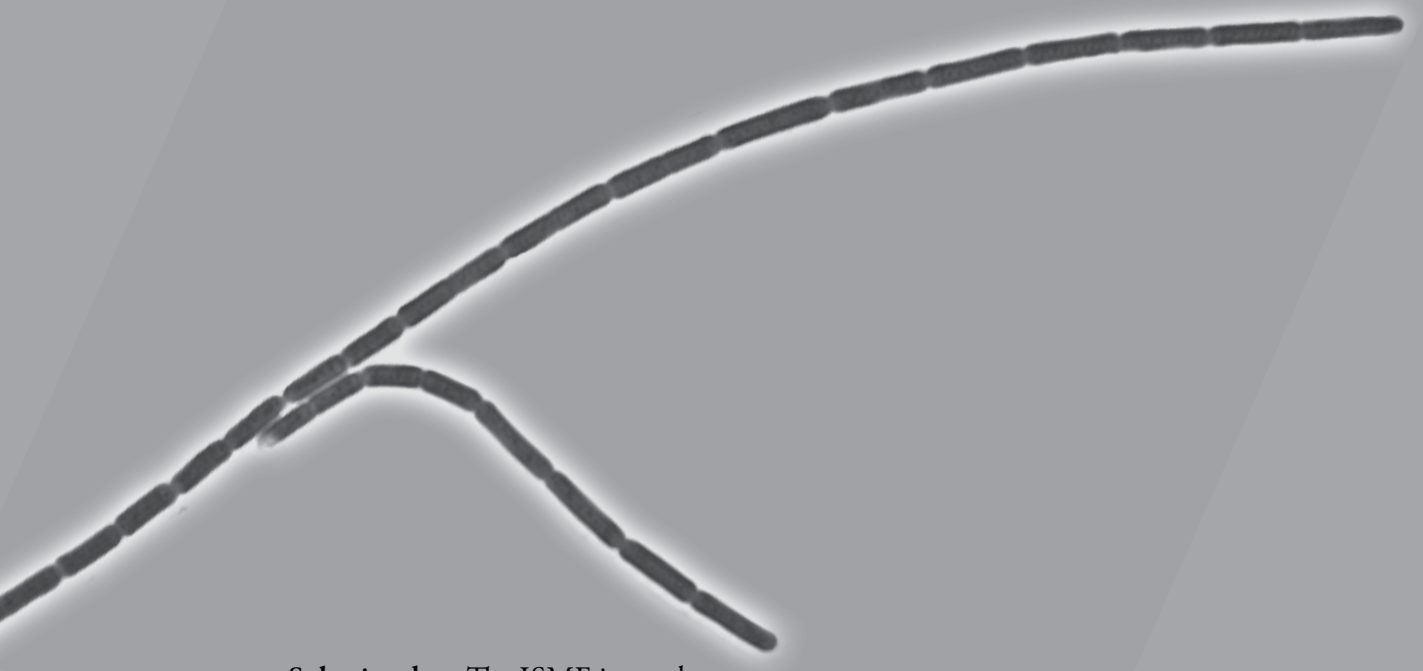
Publication date
2008

Chapter 4

Rapid diversification of red and green Synechococcus *strains in the Baltic Sea*

Thomas H.A. Haverkamp[1]
Daphne Schouten[1]
Marije Doeleman[1]
Ute Wollenzien[1]
Jef Huisman[2]
Lucas J. Stal[1,2*]

[1]*Department of Marine Microbiology, Netherlands Institute of Ecology (NIOO-KNAW), PO Box 140, 4400 AC Yerseke, The Netherlands.*
[2]*Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, The Netherlands.*



Submitted to: The ISME journal.
Running title: Microdiversity among Baltic Sea picocyanobacteria
Keywords: Baltic Sea; biogeography; horizontal gene transfer; microdiversity; phycocyanin; *Synechococcus*
Subject Category: Microbial population and community ecology

Abstract

Synechococcus is a cosmopolitan genus of picocyanobacteria living in the photic zone of freshwater and marine aquatic environments. Here, we describe the isolation of 46 closely related picocyanobacterial strains from the Baltic Sea. The isolates show considerable variation in their pigmentation phenotypes. Furthermore, small differences between the strains were observed with respect to cell volume and preference for either ammonium or nitrate as the main source of nitrogen. At the molecular level we found that 39 strains, designated BSea1, had almost identical 16S rRNA – ITS sequences with Synechococcus strain WH5701. Despite having similar 16S rRNA – ITS sequences, the BSea1 strains could be separated into several different clusters when comparing the phycocyanin (cpcBA) operon. This separation corresponds to the pigmentation of the different BSea1 strains. The majority of phycocyanin (PC) rich Bsea1 strains clustered with WH5701. Remarkably, the phycoerythrin (PE) rich strains of BSea1 formed an as yet unidentified cluster within the cpcBA phylogeny, distantly related to other PE-rich groups. Detailed analysis of the cpcBA operon using neighbour net analysis indicates that the PE-rich BSea1 strains are probably endemic for the Baltic Sea. Comparison of the phylogenies obtained by using the 16S rRNA–ITS, the cpcBA as well as the concatenated 16S rRNA-ITS and cpcBA operon sequences, revealed possible events of horizontal gene transfer among different Synechococcus lineages. Our results show that microdiversity is important in Synechococcus populations, and that it can reflect different phenotypes with different ecological roles.

Introduction

The picocyanobacterium *Synechococcus* is a typical cosmopolitan genus (Schmidt *et al.*, 1991). *Synechococcus* spp. can be found in a wide range of marine and freshwater ecosystems around the world, as well as in benthic systems and even in extreme environments such as hot springs (Schmidt *et al.*, 1991; Ramsing *et al.*, 2000; Crosbie *et al.*, 2003; Ernst *et al.*, 2003; Fuller *et al.*, 2003; Ivanikova *et al.*, 2007). *Synechococcus* comprises both picocyanobacteria (by definition <2 μm) as well as species that are larger. The genus is polyphyletic and genetically highly diverse. *Synechococcus* species are phylogenetically divided in several major clusters (Herdman *et al.*, 2001). The picocyanobacteria often found and isolated from marine, brackish and freshwaters are related to *Synechococcus* cluster 5 (Crosbie *et al.*, 2003; Ernst *et al.*, 2003). At the 16S rRNA level this cluster is closely related to the sister groups *Prochlorococcus* and *Cyanobium*. The ribosomal genes of these species form a monophyletic clade in the cyanobacterial phylogenetic tree (Herdman *et al.*, 2001; Ernst *et al.*, 2003; Fuller *et al.*, 2003).

The study of the diversity of *Synechococcus* and its sister groups has been mainly performed by analysis of the 16S rRNA gene obtained from environmental clone libraries or from isolated strains from various environments throughout the world. However, the 16S rRNA gene is conserved and has limited resolution. Several authors have therefore used additional genetic markers to study the diversity of picocyanobacteria. Examples of less conserved genetic markers used in cyanobacterial studies include the internal transcribed spacer (ITS) region of the ribosomal operon (Rocap *et al.*, 2002; Ernst *et al.*, 2003; Chen *et al.*, 2006; Haverkamp *et al.*, 2008); the phycocyanin operon (*cpcBA*) (Robertson *et al.*, 2001; Crosbie *et al.*, 2003; Ivanikova *et al.*, 2007; Haverkamp *et al.*, 2008); the phycoerythrin operon (*cpeBA*) (Everroad and Wood, 2006; Haverkamp *et al.*, 2008); the RNA polymerase core subunit gene (*rpoC1*) (Mühling *et al.*, 2006); the gene encoding for protein D1 of photosystem II (*psbA*) (Zeidner *et al.*, 2003); the RubisCO gene (*rbcL*) (Chen *et al.*, 2004; Everroad and Wood, 2006; Wilhelm *et al.*, 2006); the gene encoding the GG-phosphate synthetase (*ggpS*) and the *isiA* gene (encoding a chlorophyll *a* binding protein) (Geiss *et al.*, 2004). These studies used either isolated strains or clone libraries derived from environmental DNA. An important advantage of using cultivated strains is that it is possible to carry out phylogenetic comparisons using multiple different genetic markers. This is nowadays current practice for the analysis of pathogenic organisms in order to study the genetics behind the virulence of different strains (Hold *et al.*, 2001; Li *et al.*, 2006; Vassileva *et al.*, 2006). This approach has been rarely applied in the study of picocyanobacteria (Crosbie *et al.*, 2003; Ernst *et al.*, 2003; Chen *et al.*, 2006; Everroad and Wood, 2006). Everroad and Wood (2006) concatenated the 16S rRNA with the *rpoC1* gene in order to identify the phylogenetic position of their isolates. Crosbie *et al.* (2003) compared the phylogenies of the 16S rRNA and the *cpcBA*, Ernst *et al.* (2003) studied the 16S rRNA and the ITS, and Chen *et al.* (2006) investigated the relationship between the ITS and the *rbcL* genes.

The *Synechococcus* genus consists of both red strains rich in the pigment phycoerythrin, and green strains rich in phycocyanin. Competition experiments demonstrated that these red and green strains can coexist in white light, through niche differentiation in the light spectrum (Stomp *et al.* 2004). This spectral niche differentiation also seems to explain the distribution of

red and green strains in natural waters. A recent field survey of 70 aquatic ecosystems showed that red strains dominate in clear ocean waters, whereas green strains dominate in turbid inlands waters (Stomp *et al.* 2007a). Widespread coexistence of red and green picocyanobacteria was found in waters of intermediate turbidity, such as mesotrophic lakes and coastal seas. Based on these previous studies (Stomp *et al.* 2004, 2007a, 2007b; Haverkamp *et al.* 2008), we hypothesize that waters of intermediate turbidity, such as the Baltic Sea, are hotspots for the diversification of red and green strains.

In this study the partial 16S rRNA, the ITS region and the *cpcBA* operon sequences were concatenated in order to study the phylogenetic relationships of strains isolated from the Baltic Sea with other known picocyanobacteria. The datasets were analysed phylogenetically and by using rarefaction. We show that the concatenated dataset yielded higher estimates of cyanobacterial diversity than analysis of the datasets separately. Furthermore, we show that the multilocus approach hinted to geographical endemism and horizontal gene transfer, an observation that hints at rapid diversification of the *Synechococcus* strains in the Baltic Sea.

Materials and Methods

Sample Collection

Water samples were collected during a research cruise with the Finnish RV *Aranda* in the Gulf of Finland (Baltic Sea) from 12 - 19 July 2004. Samples were taken using a rosette sampler with Niskin Bottles. Eight stations (between N 59.1 E 22.2 and N60.0 E 26.2) (Table SM1) were sampled. A Seabird 911 CTD was connected to the rosette sampler that recorded temperature, salinity, density and oxygen concentration in the water column (Figure SM1). Water samples were taken for analysis of nutrients and processed using standard methods (Grasshoff *et al.*, 1983).

Isolation of Baltic Sea picocyanobacteria

For isolation of picocyanobacteria water samples were fractionated using a 20 μm mesh plankton net. The filtrate was used for inoculation. The filtrate was serially diluted in 96 deep well microtiterplates (Nunc, Roskilde, Denmark) containing BSea6 medium (mixture of 1 part ASNIII and 4 parts BG11 with a final salinity of 6.0 psu) (Rippka *et al.*, 1979). Instead of $NaNO_3$, NH_4Cl was used as nitrogen source (1 mM). Incubation with 10 μmol photons m^{-2} s^{-1} of light was started while still on board the ship. During transport of the enrichments to the laboratory the cultures were in the dark (for about 36 hours). Upon returning to the laboratory incubation was continued for 2 weeks at 20 °C and 10 μmol photons m^{-2} s^{-1}. Subsequently, the light level was increased to 20 μmol photons m^{-2} s^{-1} and the incubation was continued for 6 weeks until colonies became visible at the bottom of the wells. Colonies were picked with sterile glass capillaries and transferred to sterile BSea6 medium and incubated at 20 μmol photons m^{-2} s^{-1}. Clonal isolates were obtained by using the soft-agarose plating technique (Brahamsha, 1996). Briefly, dilution series were made in 96 well plates (type 655180, Greiner

Bio-one, Frickenhausen, Germany) filled with 200 µl BSea6 medium. From each dilution step 50 µl was used to inoculate 25 ml of BSea6 medium containing 0.15% (wt/vol) agarose (Sigma-Aldrich, St. Louis, MO, USA) at 40°C. After inoculation the agarose plates were allowed to solidify and left to dry. Plates were incubated under low-light (10 µmol photons m^{-2} s^{-1}) for two weeks after which the light intensity was increased to 20 µmol photons m^{-2} s^{-1} and incubation was continued until colonies appeared that were sufficiently large to be picked. Colonies were transferred to sterile BSea6 medium. After three cycles of soft-agarose plating and subsequent picking of colonies, strains were considered to be axenic and clonal and cultured in liquid BSea6 medium for experimental purposes. Strains were maintained in BSea6 amended with $NaNO_3$ (1.8 mM) and $NaHCO_3$ (0.012 mM).

Morphology and microscopy

Exponentially growing or early stationary phase cultures were collected and fixed in a mixture of 1% formaldehyde and 0.05% glutaraldehyde. The samples were stored at -80 °C until analysis. Microscope slides were prepared by covering them with a thin layer of 1% hot agarose (50 °C) that was subsequently allowed to solidify. Cells were applied to the agarose and covered with a cover slip. Observations were performed using a Zeiss Axiophot microscope (Oberkochen, Germany). Microphotographs were taken with a ProgRes C10 plus digital imaging system (JENOPTIK Laser, Optik, Systeme GmbH, Jena, Germany). The pictures were processed using ProgRes CapturePro2.0 software (JENOPTIK Laser, Optik, Systeme GmbH). At least 40 measurements of cell width and length were taken and averaged for each culture. The cell volume was calculated using the formula for a cylinder with hemispherical ends for all cells ($V = \pi (L-W) W^2/4 + \pi W^3/6$) (Sieracki *et al.*, 1989).

In vivo absorption spectra

Cultures in the exponential to early stationary phase were used for recording *in vivo* absorption spectra. An aliquot of the culture was transferred to a cuvet (4ml) and the *in vivo* absorption spectrum was measured from 400 – 750 nm using a Varian Cary 100 Bio equipped with an integrating sphere DRA-CA-3300 (Palo Alto, CA, USA).

Growth characteristics of picocyanobacterial isolates

The strains CCY0441 - CCY0492 were tested for growth with different nitrogen sources using BSea6 medium amended with $NaHCO_3$ (0.012 mM). One treatment only contained $NaNO_3$ (1.7 mM) as nitrogen source while the second treatment contained both $NaNO_3$ (1.7 mM) and NH_4Cl (1 mM) as nitrogen sources. Strains were incubated in duplicate for 3 weeks in 24-well microtiterplates (type 662160, Greiner Bio-one). After incubation, absorption spectra were measured for each strain and treatment.

DNA extraction

Stationary phase cultures were used for DNA extraction. One ml of culture was transferred to an Eppendorf tube provided by the UltraClean Soil DNA extraction kit (MoBio Laboratories

Inc., Carlsbad, CA, USA). DNA extraction was performed following the instructions of the manufacturer, except that the initial vortexing step was decreased to 5 min at maximum speed. DNA was eluted with 50 µl of TE-buffer. The DNA concentration was measured using a ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

PCR reactions

PCR reactions were performed with 20-100 ng of DNA in a 25 µl reaction volume. The 16S-ITS-operon was amplified using the primers B1055F and PitsE-cyanR (Ernst *et al.*, 2003; Zaballos *et al.*, 2006), while in a separate reaction the *cpcBA*-operon was amplified using the primers cpcBF (UFP) / cpcAR (URP) (Robertson *et al.*, 2001). For each PCR reaction the mixture contained 0.625 units of HotStarTaq (Qiagen, Venlo, The Netherlands), 0.2mM of each dNTP (Roche, Woerden, The Netherlands), 1 x PCR buffer (Qiagen), forward and reverse primers were used at a final concentration of 5 pmol and MQ-grade H_2O was added to a final volume of 25 µl.

A GeneAmp PCR System 2700 (Applied Biosystems, Foster City, CA, USA) was used to perform the PCR reactions. The PCR program to amplify the 16S rRNA-ITS-1 region consisted of a hot start at 94°C of 15 min, followed by 35 cycles of 1 min at 94°C, 1 min at 62°C and 1 min at 72°C and a final elongation step of 10 min at 72°C. The PCR of the *cpcBA*-operon was identical except that the annealing step was at 55°C and the number of cycles was 40. PCR products were checked on a 1% agarose (Sigma Aldrich) gel.

Cloning and sequencing

PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen, Breda, The Netherlands) following the instructions of the manufacturer. After overnight incubation of the transformed cells 5 colonies were picked for each PCR product and inoculated in microtiterplates (type 655180, Greiner Bio-one) containing 200 µl of liquid LB-medium. After overnight incubation at 37°C, 25 µl of culture was mixed with 25 µl of sterile MQ grade H_2O. The mixture was heated for 10 min at 95°C and subsequently used as template in a PCR reaction containing the T7 and T3 primers of the vector in order to amplify the inserted gene fragments. The PCR reaction mixture contained 0.625 units HotStarTaq (Qiagen), 0.2 mM of each DNTP (Roche), 1 x PCR Buffer (Qiagen), 5 pmol of each of the T7 and T3 primers, 2 µl of the bacterial suspension. Sterile MQ-grade H_2O was added to a final volume of 50 µl. The PCR program was run on a GeneAmp PCR System 2700 with the following settings: 15 min hot start at 94°C, 35 cycles of 30 sec at 94°C, 30 sec at 55°C, 1 min at 72 ° C and a final step of 10 min elongation at 72°C. PCR products were checked on a 1% agarose (Sigma Aldrich) gel. Positive PCR reactions were purified using the DNA clean & concentrator-5 kit (Zymo Research, Orange, CA, USA) and eluted in 20 µl of sterile MQ-grade H_2O. DNA concentrations of purified PCR products were checked with the ND1000-spectrophotometer.

For each of the loci four clones were sequenced in order to minimize errors caused by the polymerase. The sequencing reactions were performed by using the Big Dye Terminator v1.1 Cycle sequencing kit (Applied Biosystems) following the instructions of the manufacturer.

Sequence products were analyzed with a 3130 Genetic Analyzer (Applied Biosystems). For each clone, the forward and reverse sequences were manually aligned in BioEdit (Hall, 1999) and the consensus sequences were checked against GenBank using BlastN and BlastP (Altschul *et al.*, 1997; McGinnis and Madden, 2004). For all strains the 16S rRNA-ITS-1 sequences and *cpcBA* operon sequences have been submitted to the GenBank databases under accession No. EU386608 – EU386699.

Phylogenetic and diversity analysis

Phylogenetic analysis was performed using alignments of the 16S rRNA-ITS-1 and the *cpcBA* operon sequences by using closely related sequences obtained from GenBank. For the 16S rRNA-ITS-1 the highly variable regions that could not be reliably aligned were removed from the analysis, except when stated otherwise. Also, the intergenic spacer region (IGS) was removed from the *cpcBA* operon sequences before analysis for reasons of reliable alignment problems. Additionally, for those strains of which both the 16S rRNA-ITS-1 and the *cpcBA* operon are known we constructed concatenated sequences using the program DAMBE (Xia and Xie, 2001). Phylogenetic tree construction was performed using the MEGA4.0 software (Tamura *et al.*, 2007). Neighbour-net network analysis was performed using the SplitsTree4 program (Huson and Bryant, 2006). Maximum likelihood analysis using the 16S rRNA-ITS-1 region and the *cpcBA* operon was performed using PAUP* (Swofford, 2003). Analysis of the diversity at the different loci between the picocyanobacterial isolates was performed using rarefaction analysis with the software DOTUR (Schloss and Handelsman, 2005). As input for rarefaction analysis we created a distance matrix of the sequence alignments of the different loci using the program DNAdist in the Phylip package using Jukes-Cantor distances and standard settings (Felsenstein, 1989). The distance matrix was subsequently analyzed with DOTUR using standard conditions with the rarefaction option enabled. The output was used to create the rarefaction curves and similarity plots.

Results

Sampling sites

In the summer of 2004 eight stations were sampled in the Gulf of Finland (Baltic Sea). The geographic coordinates, water depth and date of sampling are summarized in Table SM1 of the supplementary materials. Vertical profiles of salinity, temperature, density, and oxygen saturation at the sampling stations are provided in Figure SM1 of the supplementary materials.

The concentrations of NH_4^+, NO_3^-, NO_2^- and PO_4^{3-} at the surface and at 30 m depth are summarized in Table 4.1. At the water surface the N:P ratios were between 1 and 2, except at station 298 where the N:P ratio was 9 due to a very low phosphate concentration. At 30 m depth the N:P ratios were higher and ranged from 3.5 – 10.6. The higher N:P ratios were mainly caused by a higher concentration of NH_4^+. Station 298 was again exceptional. The N:P ratio at

Table 4.1. *Concentrations of $PO4^{3-}$, NO_3^-, NO_2^-, and NH_4^+ (in µM) measured at the sampling stations at the surface and at 30 m depth. For station 305 we lacked data and was therefore omitted from this list.*

Stations	*PO43- (µM)*		*NO3- (µM)*		*NO2- (µM)*		*NH4+ (µM)*		**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*
S301	0.15	0.61	0	0.20	0	0.20	0.23	1.78	1.53	3.57
S307	0.09	0.48	0	0.46	0.02	0.18	0.13	1.88	1.67	5.25
S310	0.09	0.60	0.01	0.46	0.04	0.29	0.14	1.58	2.11	3.88
S314	0.09	0.32	0	0.32	0	0.08	0.11	1.56	1.22	6.13
S322	0.10	0.36	0.03	0.36	0.01	0.12	0.11	1.48	1.50	5.44
S327	0	0.18	0	0.25	0	0.01	0.06	1.65	n.d.	10.61

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

20 m was only 0.68 and was caused by low NH_4^+ and relatively high phosphate concentrations. The nutrient concentrations at this station were lower than at the other stations and the N:P ratio seemed to be largely influenced by the dynamics of phosphate (Table 4.1).

Morphology and pigmentation Forty-six picocyanobacteria were isolated from water samples of the Gulf of Finland. All of the isolates could be assigned to the picoplankton fraction based on their size and their coccoid or rod-like shape. All strains were maintained in the same cultivation medium BSea6 containing 1 mM NH_4Cl and 1.8 mM $NaNO_3$ as nitrogen sources and 0.012 mM $NaHCO_3$. Comparison of the strains showed that cell sizes and volumes varied considerably between the isolates. The list of strains, their CCY number, clade assignment, station and depth of isolation, pigmentation and size and volumes are listed in Table 4.2. The coccoid isolate CCY0452 was the smallest among the isolates (bio-volume of 0.66 μm^3) and the rod CCY0454 was the largest (bio-volume of 2.43 μm^3). Cell width varied only little between strains (0.8 - 1.4 µm), while cell length varied from 1.2 - 2.9 µm causing the bio-volume differences. Thirty-three strains contained high contents of phycoerythrin but only a little phycocyanin (PE-rich strains). The other 13 strains lacked phycoerythrin and contained larger amounts of phycocyanin (PC-rich strains) (Table 4.2). We observed that the relative absorption spectra of the strains CCY0441 - CCY0492, grown in Bsea6 medium with $NaNO_3$ and NH_4Cl, showed considerable phenotypic variation with respect to the phycobiliprotein (either PE or PC) to Chl*a* ratio, despite the fact that all strains were maintained under identical conditions (Figure 4.1; Table 4.3; Figure SM2). Furthermore, when strains CCY0441 - CCY0492 were grown in Bsea6 medium with only 1 mM $NaNO_3$ as source of nitrogen, the PE:Chl*a* or PC:Chl*a* ratios decreased in most strains (Table 4.3). All PC-rich strains showed a lower PC:Chl*a* ratio when grown with NO_3^- (Table 4.3). Many of the PE-rich strains showed a decrease in the PE:Chl*a* ratio as well, although not as strong as the PC-rich strains. The PE-rich strain CCY0456 did not show a ratio change, while the

Table 4.2. *Picocyanobacteria isolated in this study. Clade designation is based the highest similarity between the 16S rRNA-ITS sequence with known sequences found in GenBank. Cell length and width are averages of 40 cell measurements. Bio-volume of the cells was calculated from the average values of length and width using the formula for a cylinder with hemispherical ends (Sieracki et al., 1989). Strains indicated by the pigmentation PE contained high contents of phycoerythrin and only a little phycocyanin, while strains indicated by PC contained phycocyanin only.*

Strain	Clade	Site of Isolation	Station	Isolation depth	cell length (μm)	cell width (μm)	cell bio-volume (μm³)	Pigmentation
CCY0415	Subalpine C II	59° 30' N, 22° 40' E	S314	0	1.8	1.2	1.47	PC
CCY0416	Subalpine C II	59° 30' N, 22° 40' E	S314	0	nd	nd	nd	PC
CCY0417	Subalpine C II	59° 30' N, 22° 40' E	S314	0	1.5	1.1	1.07	PC
CCY0418	Subalpine C II	59° 30' N, 22° 40' E	S314	0	2.1	1.1	1.74	PE
CCY0419	Subalpine C II	59° 30' N, 22° 40' E	S314	0	1.6	1.1	1.21	PE
CCY0420	Bornholm Sea	59° 30' N, 22° 40' E	S314	0	nd	nd	nd	PE
CCY0421	Subalpine C II	59° 30' N, 22° 40' E	S314	15	1.3	1.1	0.89	PE
CCY0422	Subalpine C II	59° 30' N, 22° 40' E	S314	15	1.6	1.1	1.11	PE
CCY0423	Subalpine C II	59° 13' N, 22° 19' E	S327	15	1.9	1.1	1.48	PE
CCY0424	Subalpine C II	59° 13' N, 22° 19' E	S327	15	1.5	1.1	1.05	PE
CCY0426	Group A Cyanobium	59° 30' N, 22° 43' E	S310	0	2.0	1.3	1.90	PC
CCY0431	Subalpine C II	59° 30' N, 22° 40' E	S314	15	1.7	1.1	1.26	PE
CCY0432	Group I	60° 04' N, 26° 21' E	S298	0	nd	nd	nd	PE
CCY0434	Subalpine C II	59° 30' N, 22° 40' E	S314	0	1.5	1.2	1.20	PE
CCY0435	Subalpine C II	59° 30' N, 22° 40' E	S314	0	1.8	1.1	1.34	PE
CCY0436	Subalpine C II	59° 24' N, 22° 26' E	S322	0	1.6	1.1	1.23	PE
CCY0437	Subalpine C II	59° 13' N, 22° 19' E	S327	0	1.7	1.2	1.50	PC
CCY0439	Subalpine C II	59° 30' N, 22° 40' E	S314	0	1.6	1.1	1.17	PE
CCY0441	Subalpine C II	59° 28' N, 22° 39' E	S305	0	1.8	1.1	1.25	PC
CCY0443	Subalpine C II	59° 28' N, 22° 39' E	S305	0	2.0	1.0	1.21	PC
CCY0444	Subalpine C II	59° 30' N, 22° 50' E	S307	0	2.1	1.1	1.57	PE
CCY0446	Subalpine C II	59° 30' N, 22° 43' E	S310	0	1.3	1.1	0.83	PE
CCY0448	Subalpine C II	59° 30' N, 22° 43' E	S310	12	1.5	1.4	1.59	PE
CCY0449	Subalpine C II	59° 30' N, 22° 43' E	S310	12	2.0	1.1	1.58	PE
CCY0450	Subalpine C II	59° 30' N, 22° 43' E	S310	12	1.6	1.0	0.97	PE
CCY0451	Subalpine C II	59° 30' N, 22° 43' E	S310	12	1.7	1.0	1.07	PE
CCY0452	Subalpine C II	59° 30' N, 22° 43' E	S310	12	1.2	1.0	0.66	PC
CCY0454	Bornholm Sea	59° 30' N, 22° 43' E	S310	0	2.9	1.1	2.43	PE
CCY0455	Subalpine C II	59° 30' N, 22° 40' E	S314	0	2.1	1.1	1.62	PC
CCY0456	Subalpine C II	59° 32' N, 22° 50' E	S301	0	1.5	1.1	1.08	PE
CCY0457	Subalpine C II	59° 30' N, 22° 40' E	S314	0	1.7	1.0	1.10	PE
CCY0458	Subalpine C II	59° 30' N, 22° 40' E	S314	0	2.0	0.9	1.17	PE
CCY0461	Subalpine C II	59° 32' N, 22° 50' E	S301	0	1.5	0.9	0.80	PE
CCY0462	Subalpine C II	59° 30' N, 22° 40' E	S314	15	1.4	1.0	0.82	PC
CCY0463	Subalpine C II	59° 30' N, 22° 40' E	S314	15	1.7	1.1	1.24	PE
CCY0464	Subalpine C II	59° 30' N, 22° 40' E	S314	15	1.6	1.2	1.25	PE
CCY0465	Subalpine C II	59° 30' N, 22° 40' E	S314	15	1.8	1.1	1.31	PE
CCY0466	Subalpine C II	59° 24' N, 22° 26' E	S322	0	1.3	1.1	0.88	PE
CCY0467	Subalpine C II	59° 24' N, 22° 26' E	S322	0	1.4	1.2	1.05	PE
CCY0468	Group I	59° 32' N, 22° 50' E	S301	0	1.8	0.8	0.83	PE
CCY0469	Subalpine C II	59° 24' N, 22° 26' E	S322	15	1.3	1.0	0.82	PC
CCY0470	Bornholm Sea	59° 28' N, 22° 39' E	S305	0	1.7	1.1	1.14	PE
CCY0489	Subalpine C II	59° 28' N, 22° 39' E	S305	0	1.7	1.0	1.18	PE
CCY0490	Group I	59° 28' N, 22° 39' E	S305	0	1.8	0.9	0.86	PC
CCY0491	Subalpine C II	59° 28' N, 22° 39' E	S305	0	1.7	1.0	1.13	PE
CCY0492	Subalpine C II	59° 28' N, 22° 39' E	S305	0	1.3	1.0	0.76	PC

strains CCY0444, CCY0454, CCY0461, CCY0468 and CCY0470 showed an increase in the PE:Chl*a* ratio (Table 4.3). Phycobilisomes can be used for nitrogen storage. Hence, these results might indicate different preferences for N-sources between the strains as has been shown for other species of freshwater *Synechococcus* (Scanlan, 2003; Ernst *et al.*, 2005).

Molecular identification of the Baltic Sea picocyanobacterial isolates.

Molecular identification of the 46 strains isolated in this study revealed representatives of several clusters of picocyanobacteria. Based on sequence comparison of the ribosomal 16S rRNA - ITS-1 region, all strains were highly similar to freshwater picocyanobacteria (Table 4.2 and Figure 4.2). The isolates belong to 4 major groups within the 16S rRNA-ITS-1 phylogeny. The 16S rRNA-ITS-1 sequence of strain CCY0426 clustered with strains belonging to group A of the freshwater picocyanobacteria (Ernst *et al.*, 2003; Chen *et al.*, 2006). The second group of isolates contains the strains CCY0420, CCY0454 and CCY0470. Their 16S rRNA-ITS-1 sequences form a monophyletic clade with CCY9201 (previously BS4) and CCY9202 (previously BS5) that were previously isolated from the Baltic Sea (Ernst *et al.*, 2003; Stomp *et al.* 2004) and were assigned as 'Bornholm Sea Group'. A third clade comprises the ITS-1 sequences of the strains CCY0432, CCY0468 and CCY0490. These three strains form

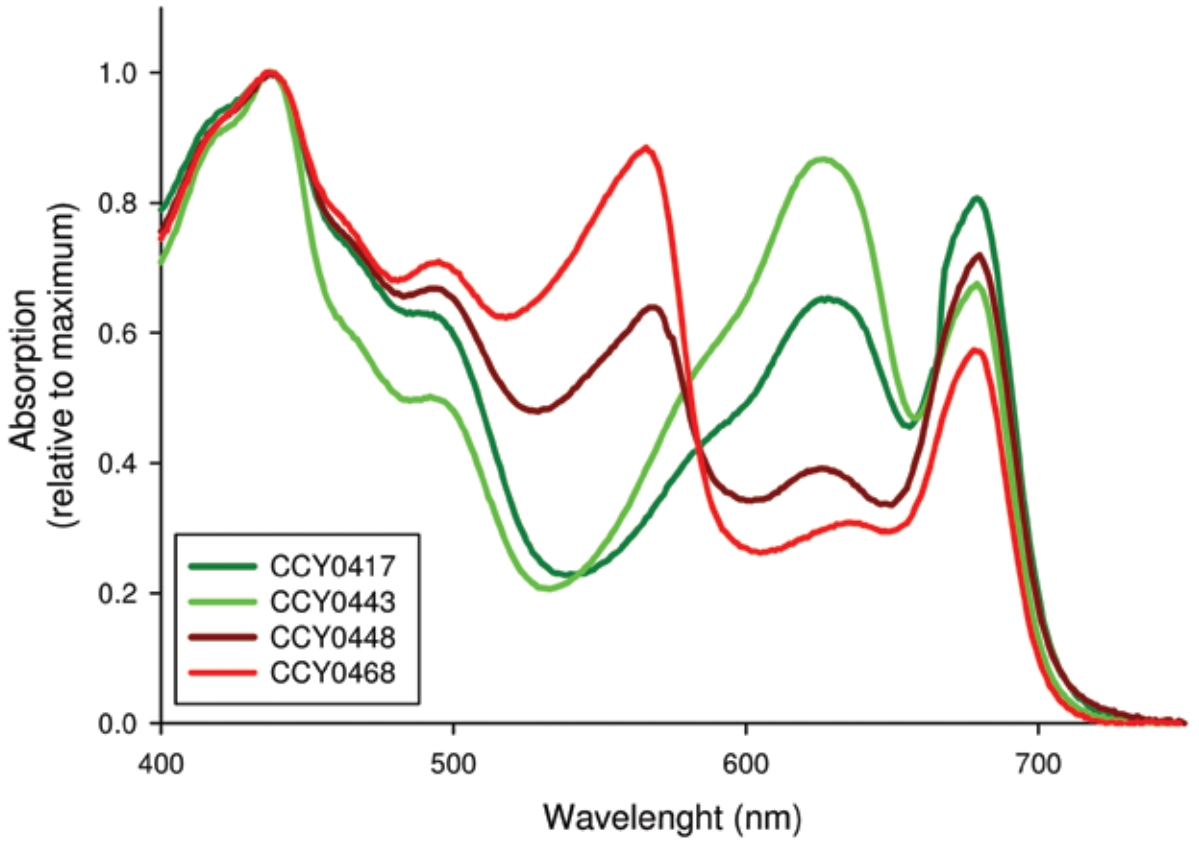


Figure 4.1. *Comparison of in vivo absorption spectra of two PE-rich and two PC-rich Baltic Sea isolates grown under the same light conditions. Note that, despite their striking difference in pigmentation, the strains CCY0417 and CCY0443 belong to the same phylogenetic clade as determined with the 16S rRNA-ITS-1 region.*

Table 4.3. *The ratio of phycobilisome absorption over Chl*a *absorption of strains CCY0441 – CCY0492, using either both $NaNO_3$ and NH_4Cl as a nitrogen source or only $NaNO_3$. For both treatments we determined the absorption peaks of the dominant phycobiliprotein (PE at 570 nm, or PC at 625 nm) and Chl*a *(680 nm), and calculated the ratio between these pigments. A ratio > 1 indicates a higher absorption by the phycobiliprotein than by Chl*a*. Absorption spectra for every strain can be found in Supplementary Figure 1.*

Strains	Pigmentation	NH_4Cl + $NaNO_3$	$NaNO_3$ Only	Observed difference
CCY0492	PC	1.2	0.5	-0.6
CCY0469	PC	1.2	0.6	-0.6
CCY0452	PC	1.1	0.6	-0.5
CCY0462	PC	1.2	0.7	-0.5
CCY0443	PC	1.1	0.7	-0.5
CCY0455	PC	1.2	0.7	-0.5
CCY0441	PC	1.2	0.9	-0.3
CCY0490	PC	1.2	1.0	-0.2
CCY0466	PE	1.2	0.7	-0.5
CCY0463	PE	1.0	0.6	-0.4
CCY0448	PE	1.0	0.6	-0.4
CCY0449	PE	1.0	0.7	-0.4
CCY0464	PE	1.0	0.7	-0.3
CCY0467	PE	1.0	0.7	-0.3
CCY0491	PE	1.0	0.7	-0.3
CCY0489	PE	1.0	0.7	-0.2
CCY0451	PE	1.2	1.0	-0.2
CCY0465	PE	0.9	0.7	-0.2
CCY0457	PE	1.0	0.8	-0.2
CCY0446	PE	1.2	1.0	-0.2
CCY0450	PE	1.2	1.0	-0.2
CCY0458	PE	1.0	0.8	-0.2
CCY0456	PE	0.9	0.9	0.0
CCY0444	PE	1.4	1.5	0.1
CCY0470	PE	1.3	1.5	0.2
CCY0468	PE	1.5	1.7	0.2
CCY0461	PE	1.5	1.9	0.3
CCY0454	PE	1.0	1.4	0.4

within the picocyanobacterial ITS-1 phylogeny, another deep branching cluster only related with the ribosomal 16S rRNA sequences of strains from group I (Crosbie *et al.*, 2003). The fourth and largest group of isolates contains 39 strains and is designated Baltic Sea group 1 (BSea1). The 16S rRNA - ITS-1 sequences of these strains show the highest similarity to those of *Synechococcus* WH5701 and BO8805 that belong to the Marine cluster B or the Sub-alpine Cluster II (Ernst *et al.*, 2003; Chen *et al.*, 2006). The 16S rRNA-ITS sequences of BSea1

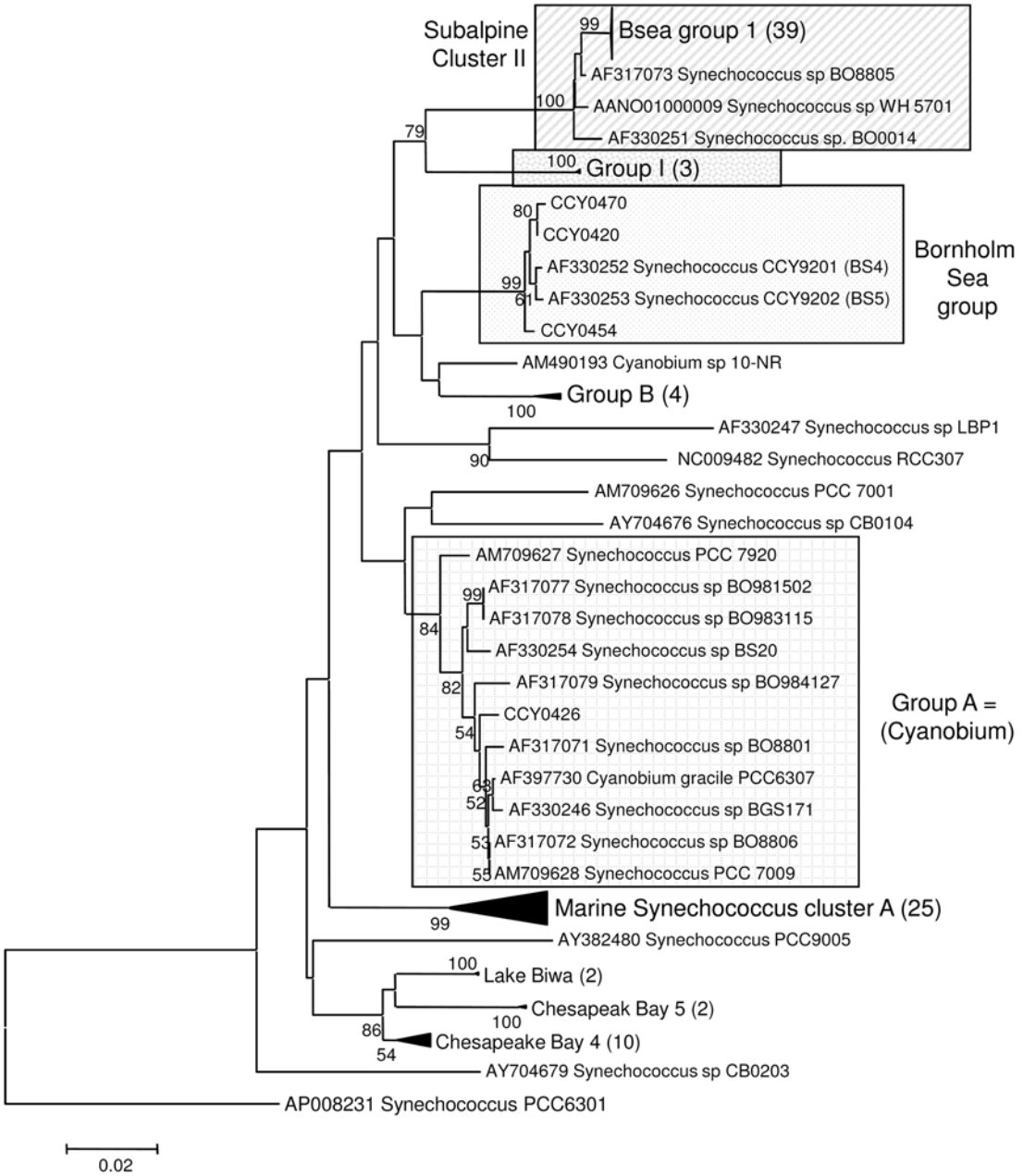


Figure 4.2. *Neighbour-joining tree based on the 16S rRNA-ITS-1 sequences from 112 picocyanobacterial isolates. The partial sequences used contained the highly conserved regions of the 16S rRNA and the ITS-1 region (862 nucleotides). The tree was rooted with the sequence of* Synechococcus *PCC6301. Construction of the Neighbour-joining tree was done using a Jukes-Cantor model of nucleotide evolution. Bootstrap values (1000 replicates) are shown for those nodes that have a confidence level higher than 50%. Taxa containing picocyanobacterial isolates from the Baltic Sea are marked by boxes. For clustered taxa it is indicated how many isolates are contained within the taxon by the numbers between brackets.*

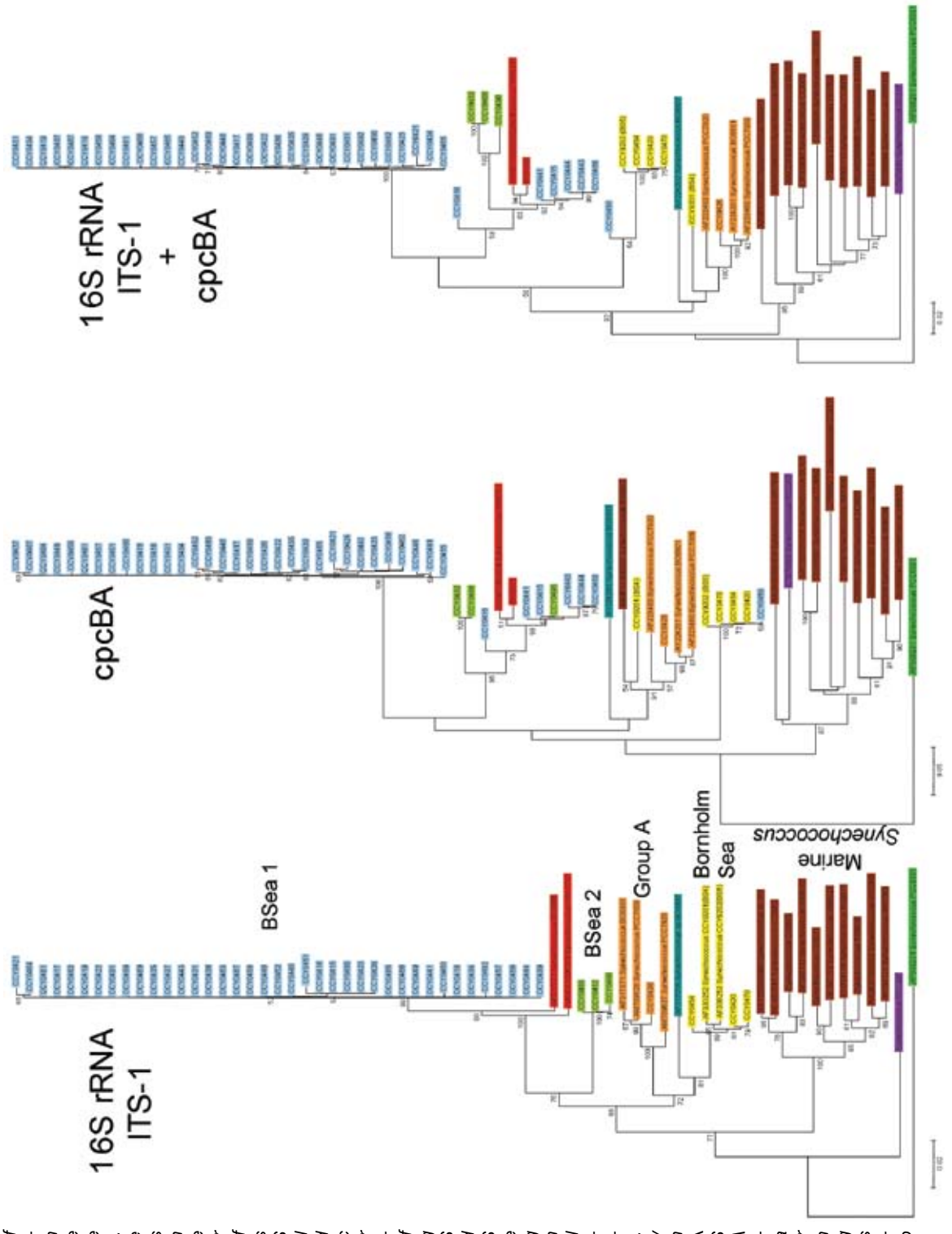


Figure 4.3. *Comparison of phylogenetic trees of the 16S rRNA-ITS region, the* cpcBA *operon and a tree based on the concatenation of the mentioned sequences. In different colours the cyanobacterial isolates that are analysed in both phylogenetic trees are indicated. Each colour represents a group of closely related strains as based on the 16S rRNA-ITS phylogeny. All trees have been created using the phylogenetic program MEGA. For the 16S rRNA-ITS phylogeny an alignment of 862 positions including the 5' end of the 16S rRNA and the conserved regions within the ITS were used. To obtain the tree a Neighbour-joining algorithm was used with the Jukes-Cantor model for nucleotide substitution and was bootstrapped 1000 times. The* cpcBA *phylogeny was obtained using an alignment of the* cpcBA *operon without the IGS (498 nucleotides). A neighbour-joining algorithm with the Kimura 2 – parameter model for nucleotide substitution was used. The resulting tree was a 1000 times bootstrapped for statistical support. See also Figure SM6.*

Table 4.4. *Comparison of diversity of picocyanobacterial* cpcBA *genes obtained in different datasets. Number of OTU is given at the similarity cut-off levels 100, 99, 97, 94 and 90%.*

Dataset	Sequences	OTU at				
		100%	99%	97%	94%	90%
Baltic Sea isolates	46	18	9	6	4	4
Baltic Sea clone libraries *	68	20	10	7	6	6
Isolates and clones	114	36	16	10	8	8
Picocyanobacterial GenBank	331	143	83	58	42	31

* *Data from Haverkamp* et al.*, 2008*

isolates can be completely aligned with WH5701 and BO8805. Nonetheless, the sequences of the BSea1 strains show on average 7.3% and 8.6% nucleotide difference at the 16S rRNA-ITS-1 region in comparison with WH5701 and BO8805, respectively. Detailed comparison of the ITS-1 of WH5701 and the BSea1 isolates revealed that most of the nucleotide differences are found in the hyper variable regions that form the hairpins (Rocap *et al.*, 2002). Within the BSea1 group, microdiversity is observed when the full length ITS-1 sequences are used for the phylogenetic analysis (Figure SM3). Four strains form a distinct cluster (BSea1B) that has on average 1.3% sequence difference with the other major sub-cluster, BSea1A (Figure SM3). *Synechococcus* WH5701 was previously assigned to *Synechococcus* cluster 5.2 and all members

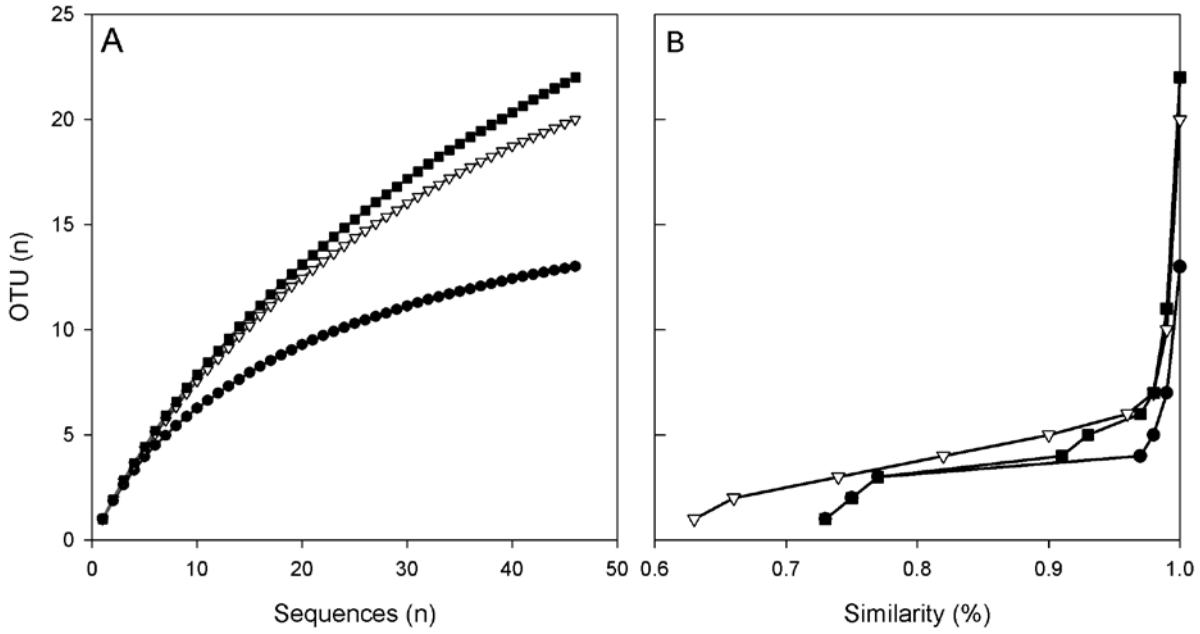


Figure 4.4. *Comparison of diversity at different loci for the Baltic Sea isolates. A) 100% similarity rarefaction curves and B) similarity plots of the 16S rRNA-ITS sequences (closed circles),* cpcBA-*operon (open triangles) and the concatenated 16S rRNA-ITS and* cpcBA *operon sequences (closed squares).*

of this clade were assumed to have phycocyanin as their main photosynthetic pigment (i.e. are green) (Herdman *et al.*, 2001). However, many of our Baltic Sea isolates are red and contain phycoerythrin as their main photosynthetic pigment.

Phycocyanin diversity of picocyanobacteria

The *cpcBA* gene found in picocyanobacteria can be used to separate strains into three different phylogenetic groups that have clearly different pigment phenotypes (Six *et al.*, 2007; Haverkamp *et al.*, 2008). In order to test whether the PE- and PC-rich Baltic Sea isolates can be separated according to pigment type, the 46 isolates were compared to known *cpcBA* operon sequences available in GenBank. The Baltic Sea PE and PC-rich isolates do cluster according to pigment phenotype as was found before (Figure SM4) (Haverkamp *et al.*, 2008). This is especially clear for the BSea1 group isolates which is split in the *cpcBA* phylogeny in a PE-rich group, clustering with other PE-rich isolates, and a PC-rich group which clusters closely with strain WH5701 and the PC-rich Group I (Figure SM4) (Crosbie *et al.*, 2003; Haverkamp

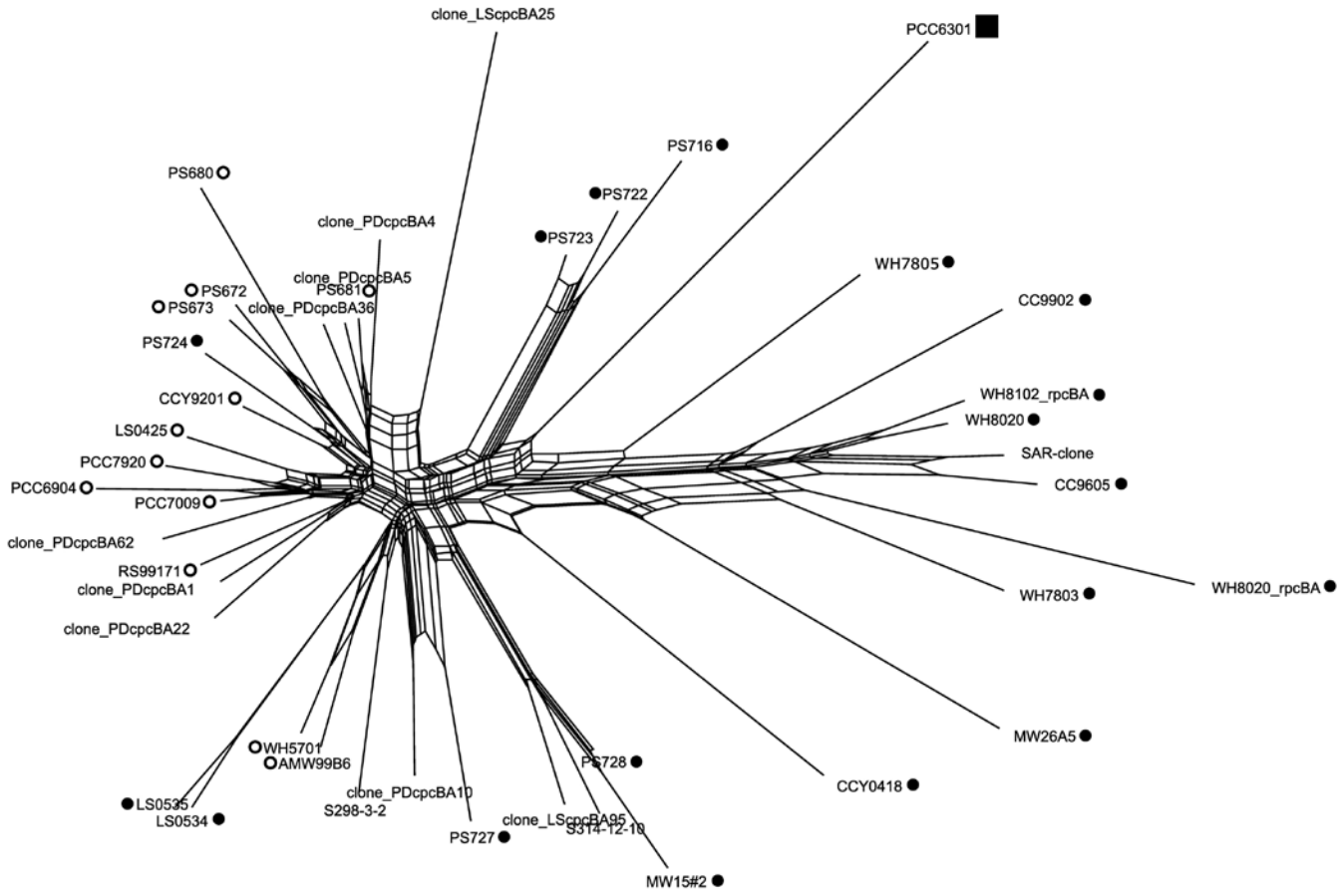


Figure 4.5. *A Neighbour-network constructed using 42 representative* cpcBA *sequences (determined by rarefaction analysis with a cut-off of 94%). Bootstrap support values for the nodes are not shown for clarity. The out-group is represented by the* cpcBA *sequence of* Synechococcus *PCC6301 (■). The PE-rich BSea1 cluster is represented by strain CCY0418. Pigmentation phenotypes are in indicated for the strains used here; PE-rich strains (•), PC-rich strains (o). For a complete analysis using all known cpcBA sequences see Figure SM7.*

et al., 2008). Nonetheless, a more detailed representation of the same phylogenetic tree shows that several PC-rich strains do cluster with the PE-rich isolates of the BSea1 group (Figure SM5). Remarkably, the PE-rich isolates CCY0432 and CCY0468 fall into the Baltic Sea cluster of clones in Group I, which until now only contained PC-rich isolates (Figure SM5). The *cpcBA* sequences of PE-rich strains CCY0420, CCY0450, CC0454 and CCY0470 form a monophyletic clade with CCY9202 and strains belonging to PE-rich group E. Finally, the *cpcBA* operon sequence of PC-rich strain CCY0426 clusters together with the PC-rich group A (Crosbie *et al.*, 2003) (Figure SM4; Figure SM5).

In addition to the study of isolates from the Baltic Sea, clone libraries were constructed from environmental DNA extracted from the same water samples (Haverkamp *et al.*, 2008). The combination of the *cpcBA* dataset from clone libraries and from the isolates revealed the highest diversity in the combined dataset (Table 4.4; Figure SM5). Phylogenetic analysis revealed that this is caused by the presence of groups that are only found in one of the data sets, thus increasing the overall diversity (Figure SM5). Rarefaction analysis of the *cpcBA* sequences of the isolates, the clone libraries and additional sequences obtained from GenBank (331 sequences in total), showed that even at the 90% similarity level, the global diversity of the *Synechococcus cpcBA* operon was not saturated (data not shown). In total 31 OTUs representing major clades within the picocyanobacteria were found using a 90% similarity cut-off (Table 4.4).

Comparison of the 16S rRNA-ITS-1 and cpcBA phylogenies

Using 16S rRNA we found a close relationship between the PC-rich strains WH5701, BO8805 and the BSea1 group (both PE- and PC-rich) of our isolates. In addition, the *cpcBA* operon suggests a split between the PE and PC-rich picocyanobacteria from the BSea1 group.

However, a clear comparison between the two phylogenetic trees is not possible since both are using different sequences which can create differences in the phylogenetic analysis. Therefore we used only *cpcBA* sequences from isolates for which also the 16S rRNA-ITS-1 sequences are available from GenBank for the phylogenetic analyses. This allowed for comparison of the phylogenies of the 16S rRNA-ITS-1 and *cpcBA* sequences of these strains and provided the possibility of concatenation of these sequences.

Most of the strains that are closely related at the 16S rRNA-ITS level were also closely related for the *cpcBA* region (Figure 4.3; Figure SM6). However, several strains have different positions within the *cpcBA* phylogeny when compared to the 16S rRNA-ITS-1 phylogeny (Figure 4.3). This is especially obvious for strain CCY0450, which belongs to the BSea1 group at the 16S rRNA-ITS-1 level, but clusters with the Bornholm Sea group based on *cpcBA* phylogeny. In the 16S rRNA-ITS phylogeny the Bornholm Sea group is clearly separated from BSea1. Other strains that behave in a similar way are CCY0416 (BSea1), CCY0490 (Bsea2), CCY9201 (Bornholm Sea) and the marine *Synechococcus* strains RS9917, and RC307 (Figure SM6) (Fuller *et al.*, 2003; Six *et al.*, 2007; Haverkamp *et al.*, 2008).

Interestingly, based on the *cpcBA* phylogeny, the BSea1 group is split into a group composed mainly of PE-rich strains and a group of PC-rich strains that clusters with the PC-rich strains WH5701 and BO8805. This contradicts the 16S rRNA-ITS-1 phylogeny in which PE-rich

strains of the BSea1 group form a monophyletic clade with PC-rich strains including WH5701 and BO8805. Hence, the phylogeny of the 16S rRNA-ITS is incongruent with that of the *cpcBA* (Figure 4.3). It is difficult to compare the phylogenetic trees of 16S rRNA-ITS and *cpcBA* because of assumptions used to create the former. In order to construct the 16S rRNA-ITS-1 phylogenetic tree large variable regions had to be removed. However, the closely related strains of the BSea1 group, WH5701, BO8805 and the less similar group I strains can be used to test if the topologies are different using Maximum Likelihood. The likelihoods of trees of 16S rRNA-ITS and the *cpcBA* operon were evaluated using the Jukes Cantor model of nucleotide substitution by fitting the tree of one gene to the ML tree of the other followed by re-sampling (Swofford, 2003). The resulting distributions of log-likelihoods and associated p-values were compared to that of the best ML topology (Table 4.5). The results of this exercise indicated that for the tested clade the 16S rRNA-ITS-1 phylogeny is significantly different from that of the *cpcBA* operon. Removal of the group I from the analysis showed no significant differences between the two phylogenies.

Finally, both datasets were used to create concatenated sequences of the 16S rRNA-ITS-1 and the *cpcBA* operon. The concatenated sequences were used to construct a phylogeny revealing a tree resembling the *cpcBA* (Figure 4.3). Most clades found in the *cpcBA* phylogeny appeared also in the concatenated dataset, showing the high diversity of *cpcBA*. Yet, in the concatenated dataset CCY0490 and RS9917 cluster with strains for which they have a closely related 16S-rRNA-ITS-1 sequence, but a different *cpcBA* sequence. Hence, concatenated 16S-rRNA-ITS and *cpcBA* sequences changes the picture only slightly when compared to the *cpcBA* alone (Figure 4.3 and Figure SM6). This conclusion is supported by rarefaction analysis of the datasets, where concatenation increased the diversity at the 100% similarity level to 22 OTU compared to 20 and 13 OTU for the *cpcBA* operon and 16S rRNA-ITS1, respectively (Figure 4.4).

Split network analysis

The *cpcBA* phylogeny is represented by well supported clades deep in the tree (Figure SM4 and SM5). However, the basal taxa of the *cpcBA* phylogeny are shown with long branches and the nodes have low bootstrap support (<50%) (Figure SM4). This could be an indication for homoplasy (backward and/or parallel mutations). Calculation of the homoplasy index HI in MEGA4.0 revealed high levels of homoplasy in the data set (HI=0.74 for parsimony-informative sites). This high value of the HI suggests that there is a large amount of ambiguity within the data that could have an effect on the phylogenetic three. A possible way to study this is by analyzing the *cpcBA* dataset using Neighbour-net networks (Bryant and Moulton, 2004). Neighbour-net networks can be used to study aligned sequence data by creating a network rather than a tree in order to show conflicting data within the phylogeny by representing the data and the evolutionary distances (Huson and Bryant, 2006). A Neighbour-net network was created using the *cpcBA* sequences (Figure 4.5; the same analysis using the complete set of sequences can be found Figure SM7). The occurrence of block shaped patterns instead of a tree like branching pattern indicates a high uncertainty in the phylogenetic signal that is probably caused by homoplasy or recombination. This view is consistent with the finding that

the phylogenies of the 16S rRNA-ITS-1 sequences and the *cpcBA* operon sequences are not entirely congruent. Nonetheless, the Neighbour-net network shows by the long branches that the *cpcBA* sequences of the PE-rich BSea1 cluster (represented by strain CCY0418) and the other groups form distinct taxa within the *cpcBA* phylogeny of the picocyanobacteria (Figure 4.5 and Figure SM7). In addition it shows a clear separation between the *cpcBA* genes of PE-rich and PC rich picocyanobacteria (Figure 4.5).

Discussion

Nitrogen preference of Baltic Sea isolates.

The isolation of 46 *Synechococcus* strains from the Baltic Sea enabled us to study the microdiversity found in this group of cyanobacteria. The strains were compared using morphological, physiological and molecular approaches. At the morphological level the main difference between isolates is that they are either PE-rich or PC-rich. This division is genetically supported using the *cpcBA* genes (Six *et al.*, 2007; Haverkamp *et al.*, 2008). Other features like cell volume and nitrogen preferences could not easily be separated into different clusters at the genetic level. Nonetheless, several strains showed a preference for nitrate over ammonium and two of these strains, CCY0454 and CCY0470, fall in a distinct cluster and are closely related to the Bornholm Sea strains CCY9201 (BS4) and CCY9202 (BS5) (Figure 4.2). The latter strains were isolated previously using a mixture of ASNIII and BG11 containing NO_3^- as the main nitrogen source (Ernst *et al.*, 2003). This suggests that isolates of the Bornholm Sea group are better adapted to nitrate than the majority of our Baltic Sea isolates who seem to prefer ammonium. The ammonium preference of the majority of our strains is logical in several ways. Ammonium was the dominant nitrogen source at our sampling stations (Table 4.1). In addition, ammonium was the nitrogen source used for isolation, thus we selected for isolation of species growing efficiently on ammonium. Furthermore, previous studies showed that nitrogen addition using NH_4^+ increased the abundance of picocyanobacteria in Baltic Sea waters to a much greater extent than nitrate (Stal *et al.*, 1999; Kuuppo *et al.*, 2003; Lagus *et al.*, 2007), suggesting that the majority of Baltic Sea picocyanobacteria have a preference of ammonium over nitrate.

The ammonium preference of our Baltic Sea isolates can be explained in the light of the recent publication of Agawin *et al.* (2007). They found coexistence between a N_2-fixing *Cyanothece* strain and a *Synechococcus* strain in continuous cultures. The low nitrate input supplied to the continuous cultures could not explain the high abundance of *Synechococcus* cells observed in the species mixture. Instead, excretion of fixed nitrogen by the *Cyanothece* strain provided enough additional nitrogen in the growth medium to allow a high abundance of *Synechococcus* cells (Agawin *et al.*, 2007). At our sampling stations, nitrate concentrations were very low and the larger phytoplankton (<20 um) consisted of large filamentous N_2-fixing cyanobacteria such as *Nodularia spumigena* (Haverkamp *et al.*, 2008). Fixed nitrogen is released by N_2-fixing cyanobacteria in the form of ammonium or dissolved organic nitrogen (DON), which could

select for a picocyanobacterial population with a higher affinity for ammonium than for nitrate (Ohlendieck *et al.*, 2007). This would explain why we found only a few isolates from the Bornholm Sea group since they seem to prefer nitrate over ammonium and thus would be out competed by species with an ammonium preference in the setup of our isolation strategy. Moreover, this suggests that strains belonging to the Bornholm Sea group are not dominant members of the picoplankton in the Baltic Sea under nitrate-limited conditions (Stal *et al.*, 2003).

Diversity of Baltic Sea Synechococcus.

The use of molecular markers to analyze phylogenetic relationships of our *Synechococcus* strains from the Baltic Sea with known isolates and clone sequences allowed us to study the diversity found in the Baltic Sea. The markers were the 16S rRNA-ITS region and the *cpcBA* operon. In a previous study these loci were used to study the diversity within clone libraries constructed from environmental DNA (Haverkamp *et al.*, 2008). Several groups of picocyanobacteria found in the clone libraries were not found among the cultured isolates, and *vice versa* (Table 4; Figure SM5). This demonstrated that neither approach sampled the full existing diversity of picocyanobacteria. Hence, clone libraries and culture collection complement each other (Kisand and Wikner, 2003; Alonso *et al.*, 2007). The combination of *cpcBA* sequences from cultures and environmental clone libraries increased the picocyanobacterial *cpcBA* sequence diversity found in the Baltic Sea considerably.

Isolated strains, however, have an advantage for the study of diversity in comparison to clone libraries, since they give the possibility to relate genotype to phenotype. In our study, strains of the BSea 1 group are closely related at the ITS level to the Marine Cluster 5.2 and Subalpine Cluster II strains WH5701 and BO8805 (Herdman *et al.*, 2001; Crosbie *et al.*, 2003; Ernst *et al.*, 2003; Becker *et al.*, 2004; Chen *et al.*, 2006). We found both PC-rich and PE-rich isolates in the BSea 1 group, all closely related at the ITS level. The different pigmentation types could not be separated using the ribosomal 16S rRNA-ITS sequences.

Conversely, analysis of the *cpcBA* operon shows that PE-rich and PC-rich strains from the BSea 1 group can be separated. This implies that although PE-rich and PC-rich *Synechococcus* strains have similar 16S rRNA-ITS-1 sequences, they have quite different pigment genotypes, enabling their adaptation to different light climates (Six *et al.*, 2007; Stomp *et al.*, 2007a; Haverkamp *et al.*, 2008). Furthermore, our closely related isolates behave quite differently towards different nitrogen sources. Therefore, quantifying 16S rRNA or ITS-1 genotypes using real-time PCR or FISH based probing in the environment while assuming certain phenotypic traits connected to these ribosomal genotypes (usually based on a few isolated strains) seems prone to and misinterpretation. This is especially difficult in the light of the extensive microdiversity found in bacteria in general (Acinas *et al.*, 2004). In our opinion, it would be better to use genes closely connected to certain phenotypes, like the *cpcBA* operon, in order to assess the ecological distribution of these different phenotypes in the environment. That would probably give more meaningful observations related to the ecological niche of bacteria.

Micro-diversity, recombination and endemism

Our results suggest that the diversity of picocyanobacteria in the Baltic Sea is substantially higher than we could measure. Rarefaction analysis using all known picocyanobacterial *cpcBA* sequences shows that the diversity found within the picocyanobacterial *cpcBA* gene is not levelling off, even not at the lower similarity levels (Table 4.4). Hence, it is not surprising that when an unexplored ecosystem, like the Baltic Sea, is sampled the first time it will yield new additional taxa, thereby increasing the global diversity of picocyanobacterial *cpcBA* sequences (Kemp and Aller, 2004). One possible explanation for the high microdiversity of *Synechococcus* strains in the Baltic Sea is that the *cpcBA* genes are quickly evolving, with the third codon position showing many synonymous mutations that increases microdiversity (Haverkamp *et al.*, 2008). Additionally, the high picocyanobacterial diversity could be caused by high recombination in the form of horizontal gene transfer. We considered the possibility of recombination within the genomes of isolated strains by comparing the phylogenies of the 16S rRNA-ITS-1, the *cpcBA operon*, and a concatenate of both sequences. A Maximum likelihood analysis of the phylogenies suggests that recombination has taken place between the different strains thereby creating higher diversity. In addition, we analyzed the *cpcBA* operon using the Neighbour-net analysis, which showed high ambiguity between the sequences. This indicated that evolutionary processes within the *cpcBA* operon do not behave in a tree-like manner. This behaviour can be explained in three ways, by extensive homoplasy (backward and/or parallel mutations), horizontal gene transfer of the *cpcBA* operon, or by a lack of data. These results are in line with genome analysis of the phycobiliprotein gene complexes of several marine *Synechococcus* strains. The genes encoding phycocyanin, phycoerythrin I and II showed different evolutionary relationships in comparison to the genes belonging to the genome core such as the allophycocyanin gene or the ribosomal 16S rRNA-ITS sequences (Six *et al.*, 2007).

Finally, the neighbour-net analysis indicated that on the basis of the *cpcBA* phylogeny the PE-rich isolates of the BSea1 group are distantly related to other *Synechococcus* isolates and in that way might be specific for the Baltic Sea with its unique environmental conditions. Studies on large ecosystems such as Lake Superior and the Baltic Sea indicate that these environments inhabit locally adapted *Synechococcus* spp. (Ivanikova *et al.*, 2007; Haverkamp *et al.*, 2008). These results indicate that the *cpcBA* operon can be used to study diversity of *Synechococcus* species from distinct environments. This is consistent with results of Pommier *et al.* (2007), who found signals of endemism within the global bacterioplankton community. In line with our previous and current findings, one likely explanation for the tremendous diversity of picocyanobacteria in the Baltic Sea is that its underwater light spectrum offers suitable niches for the coexistence of red and green picocyanobacteria, which may favour the rapid diversification of PE-rich and PC-rich strains.

Acknowledgments

We thank M. Laamanen for the opportunity to join cruise CYANO-04, M. Stomp and the crew of the research vessel Aranda for help during sampling. We also thank T. Mes for constructive discussion and help with the Maximum likelihood comparisons. T.H. and L.J.S. acknowledge support from the European Commission through the project MIRACLE (EVK3-CT-2002–00087). The research of J.H. was supported by the Earth and Life Sciences Foundation (ALW), which is subsidized by the Netherlands Organization for Scientific Research (NWO). This is publication xxxx of NIOO-KNAW.

Supplementary Information.

The following supplementary material is available for this article:

Table SM1 Overview of the location, depth and the time of sampling for all Baltic Sea Sampling stations.

Figure SM1 Depth profiles for A) salinity, B) temperature, C) density and D) oxygen saturation at the stations 298, 301, 305, 307, 310, 314, 322 and 327.

Figure SM2 Absorption spectra for all tested strains.

Figure SM3 Neighbour-joining tree using 16S rRNA-ITS-1 sequences (1045 nucleotides) from the isolates of the Baltic Sea 1 group and the closely related *Synechococcus* isolates WH5701 and BO8805.

Figure SM4 Phylogenetic tree based on the *cpcBA* operon sequences of the Baltic Sea *Synechococcus* isolates and available sequences obtained from GenBank (498 positions).

Figure SM5 Radial phylogenetic tree using *cpcBA* operon sequences of the Baltic Sea *Synechococcus* isolates and available sequences obtained from GenBank (498 positions).

Figure SM6 Comparison of phylogenetic trees of the 16S rRNA-ITS region, the *cpcBA* operon and a tree based on the concatenation of the mentioned sequences.

Figure SM7 Neighbour-net network based on the 342 *cpcBA* operon sequences.

Table SM1: *Baltic Sea sampling station, geographic location, depth of the water column and the sampling date.*

Station	Latitude	Longitude	Bottom depth (m)	Date
St298	60,04	26,21	69	120704
St301	59,32	22,50	75	140704
St305	59,28	22,39	79	150704
St307	59,30	22,50	85	150704
St310	59,30	22,44	77	160704
St314	59,30	22,40	69	170704
St322	59,24	22,26	90	180704
St327	59,13	22,19	112	180704

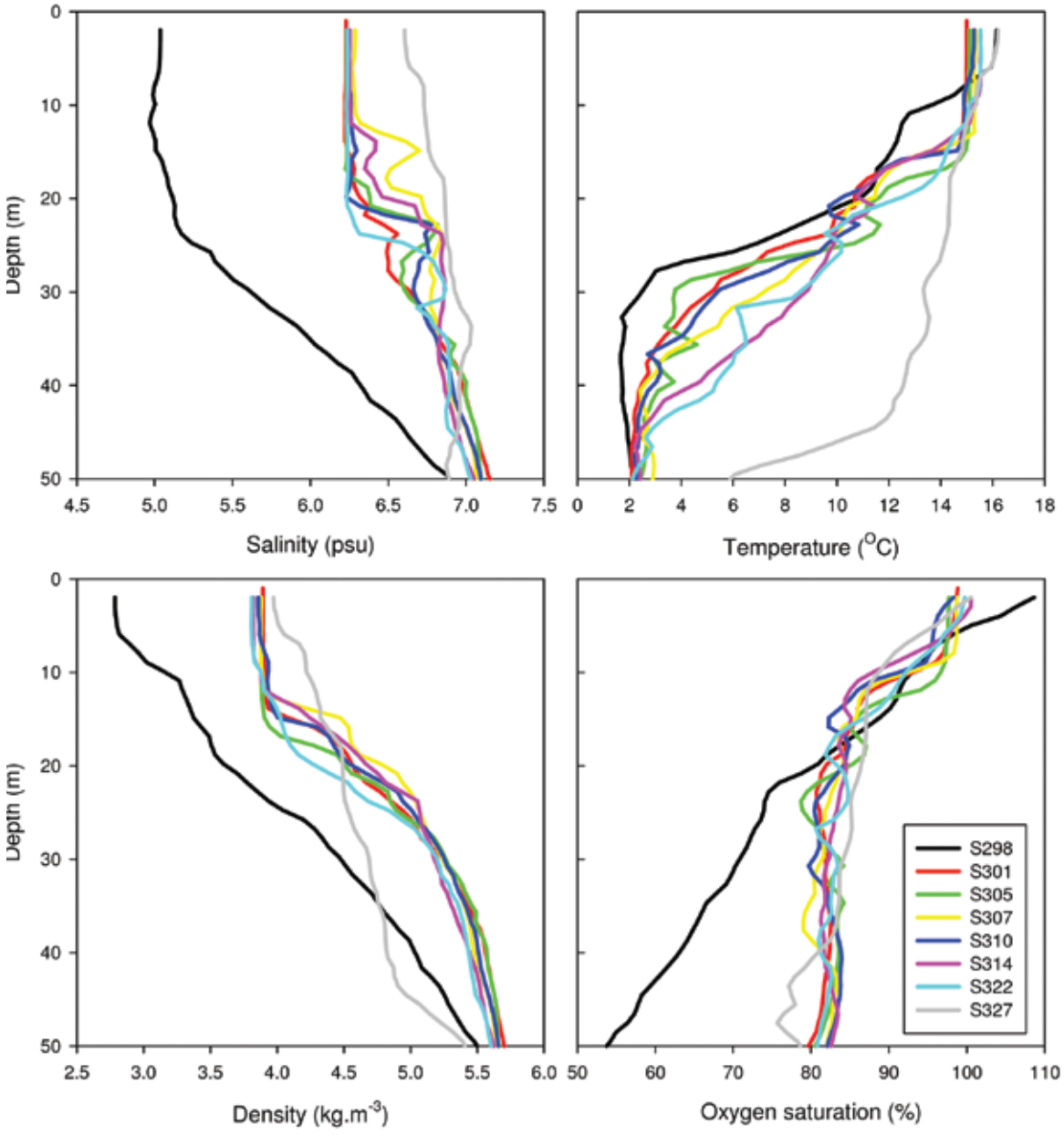

Depth (m)
Salinity (psu)
Temperature (°C)
Density (kg.m-3)
Oxygen saturation (%)
S298
S301
S305
S307
S310
S314
S322
S327

Figure SM1

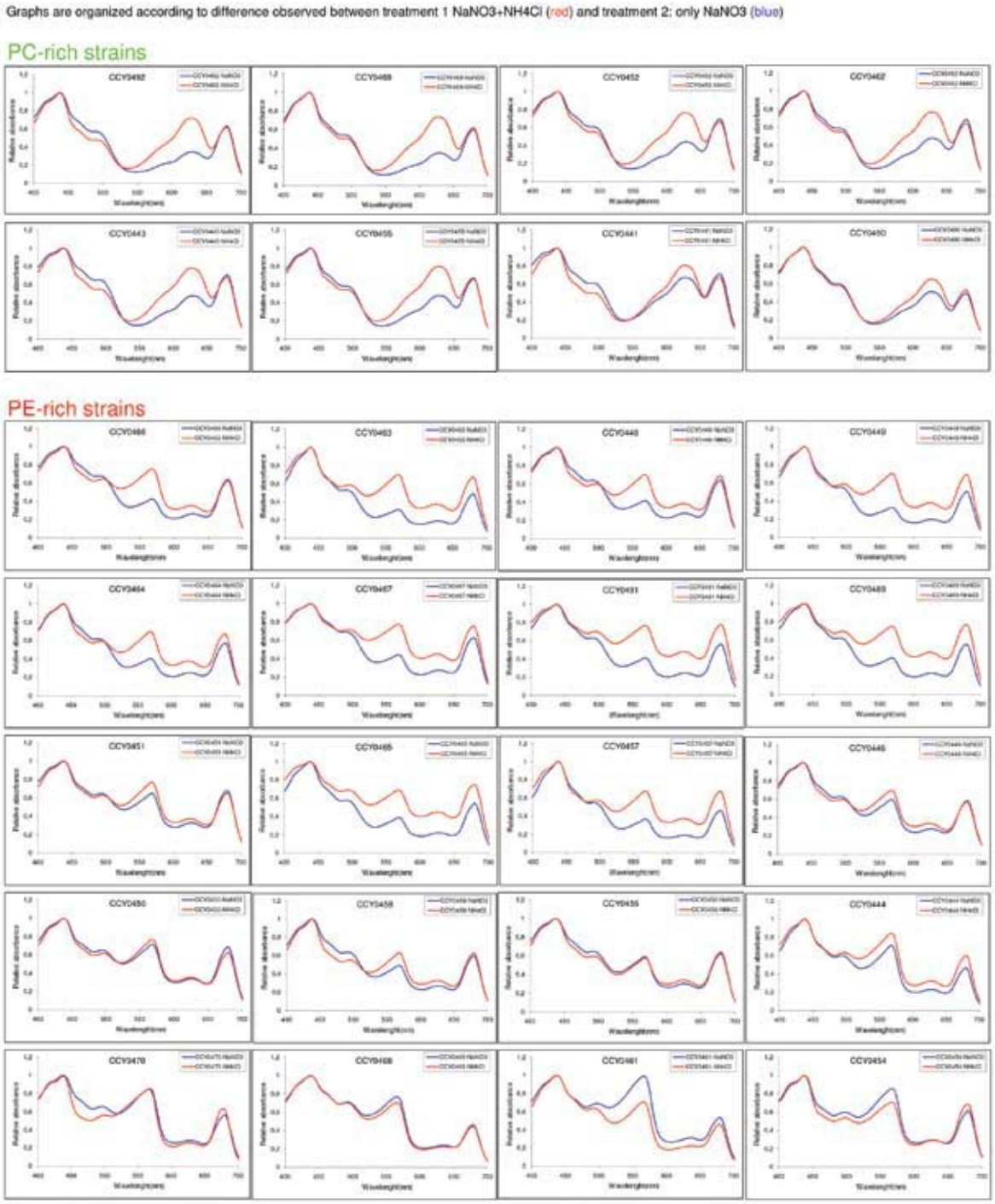

Graphs are organized according to difference observed between treatment 1 NaNO3+NH4Cl (red) and treatment 2: only NaNO3 (blue)
PC-rich strains
PE-rich strains

Figure SM2

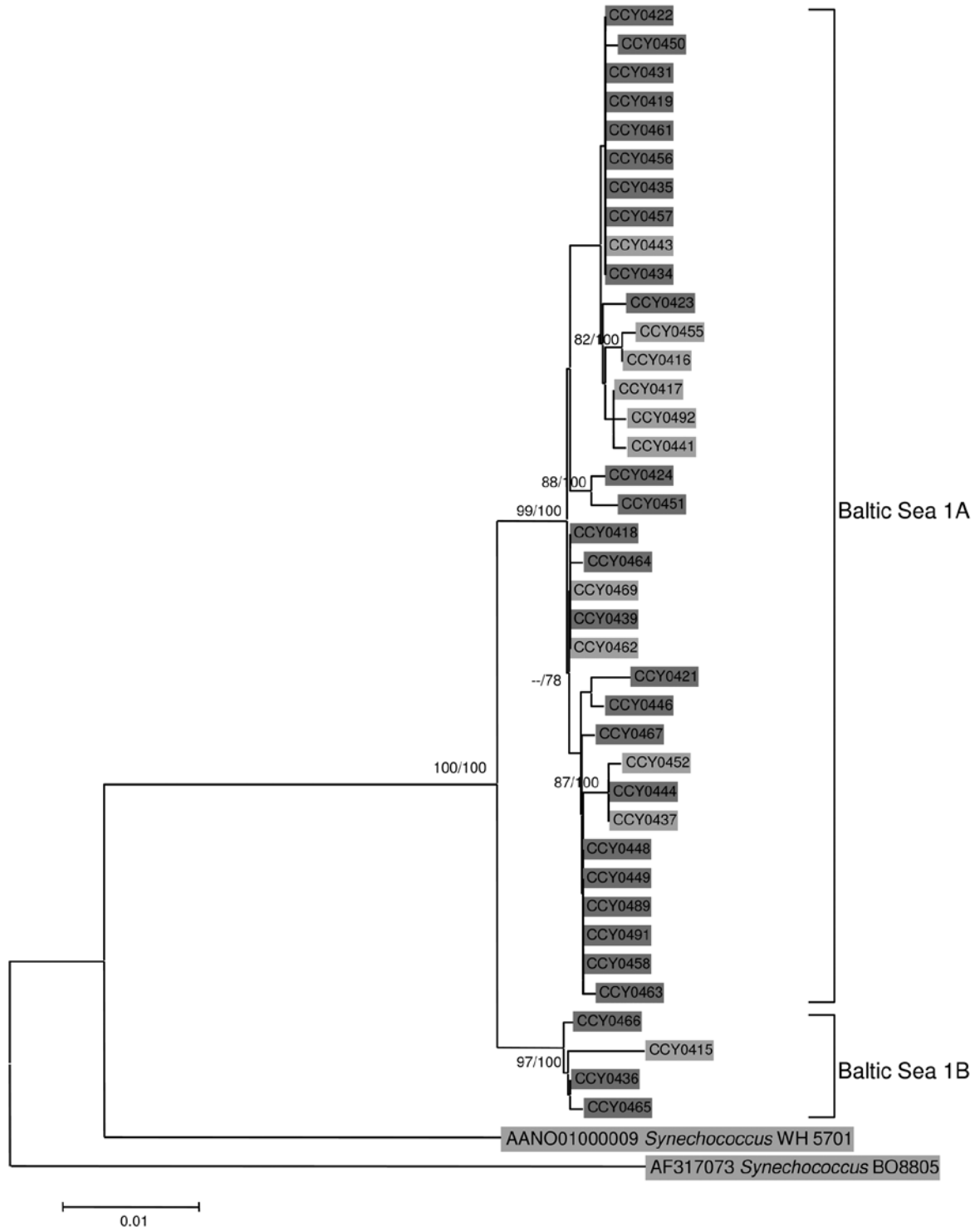

CCY0422
CCY0450
CCY0431
CCY0419
CCY0461
CCY0456
CCY0435
CCY0457
CCY0443
CCY0434
CCY0423
CCY0455
82/100
CCY0416
CCY0417
CCY0492
CCY0441
CCY0424
88/100
CCY0451
99/100
CCY0418
CCY0464
CCY0469
CCY0439
CCY0462
--/78
CCY0421
CCY0446
CCY0467
CCY0452
100/100
87/100
CCY0444
CCY0437
CCY0448
CCY0449
CCY0489
CCY0491
CCY0458
CCY0463
CCY0466
CCY0415
97/100
CCY0436
CCY0465
Baltic Sea 1A
Baltic Sea 1B
AANO01000009 Synechococcus WH 5701
AF317073 Synechococcus BO8805
0.01

Figure SM3

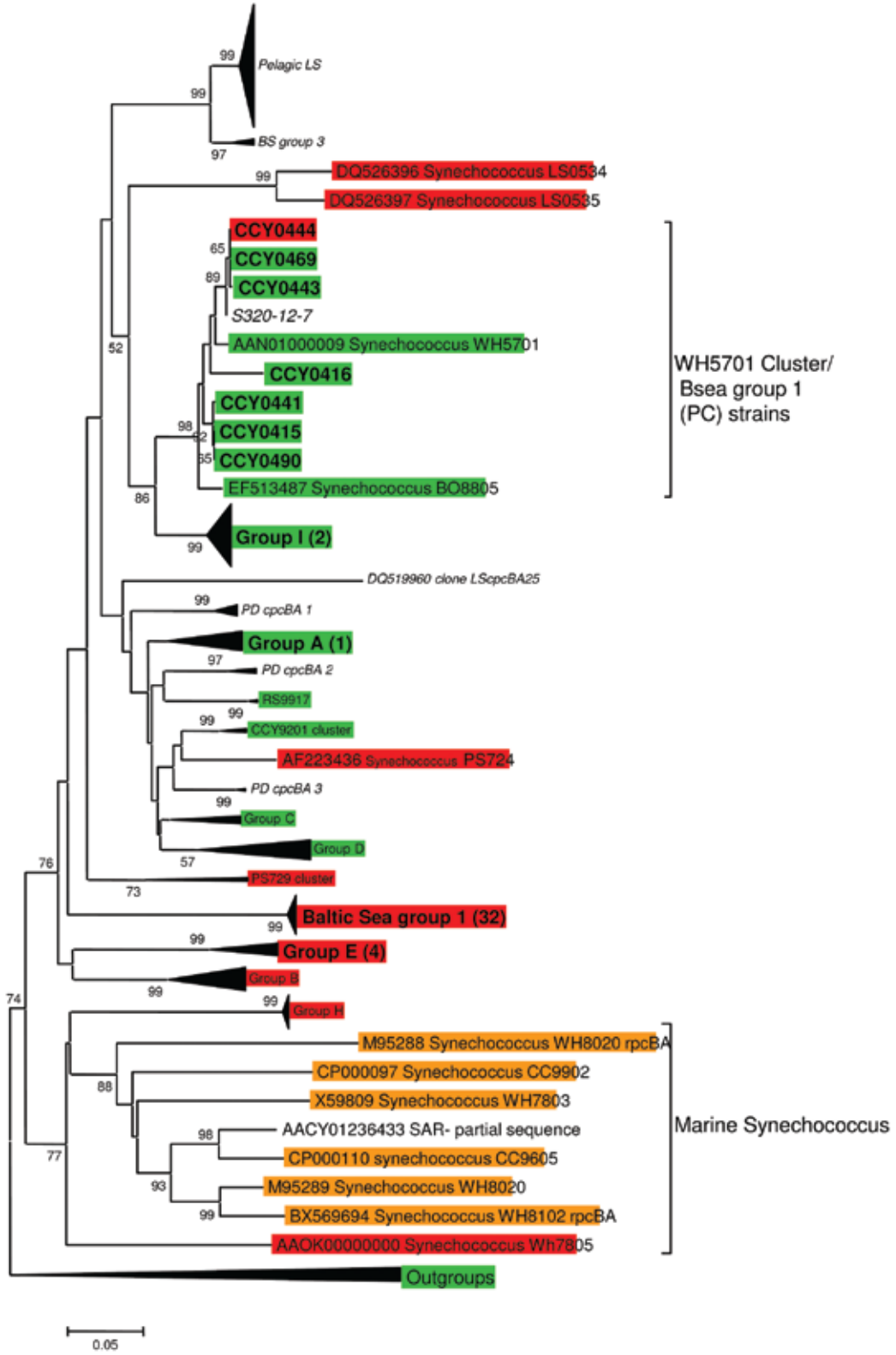

Pelagic LS
BS group 3
DQ526396 Synechococcus LS0534
DQ526397 Synechococcus LS0535
CCY0444
CCY0469
CCY0443
S320-12-7
AAN01000009 Synechococcus WH5701
CCY0416
CCY0441
CCY0415
CCY0490
EF513487 Synechococcus BO8805
WH5701 Cluster/
Bsea group 1
(PC) strains
Group I (2)
DQ519960 clone LScpcBA25
PD cpcBA 1
Group A (1)
PD cpcBA 2
IRS9917
CCY9201 cluster
AF223436 Synechococcus PS724
PD cpcBA 3
Group C
Group D
PS729 cluster
Baltic Sea group 1 (32)
Group E (4)
Group B
Group H
M95288 Synechococcus WH8020 rpcBA
CP000097 Synechococcus CC9902
X59809 Synechococcus WH7803
AACY01236433 SAR- partial sequence
CP000110 synechococcus CC9605
M95289 Synechococcus WH8020
BX569694 Synechococcus WH8102 rpcBA
AAOK00000000 Synechococcus Wh7805
Marine Synechococcus
Outgroups
0.05

Figure SM4

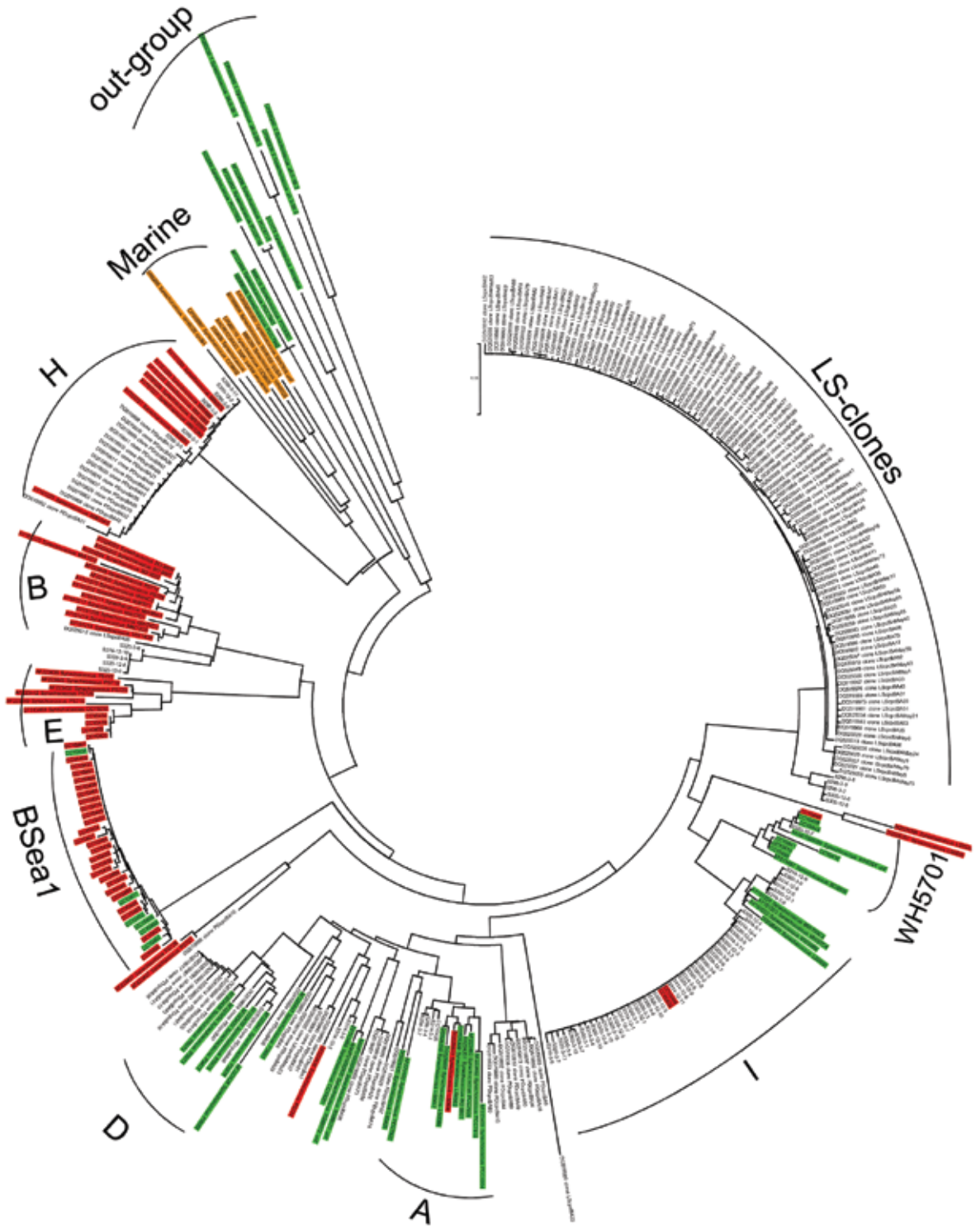


Figure SM5

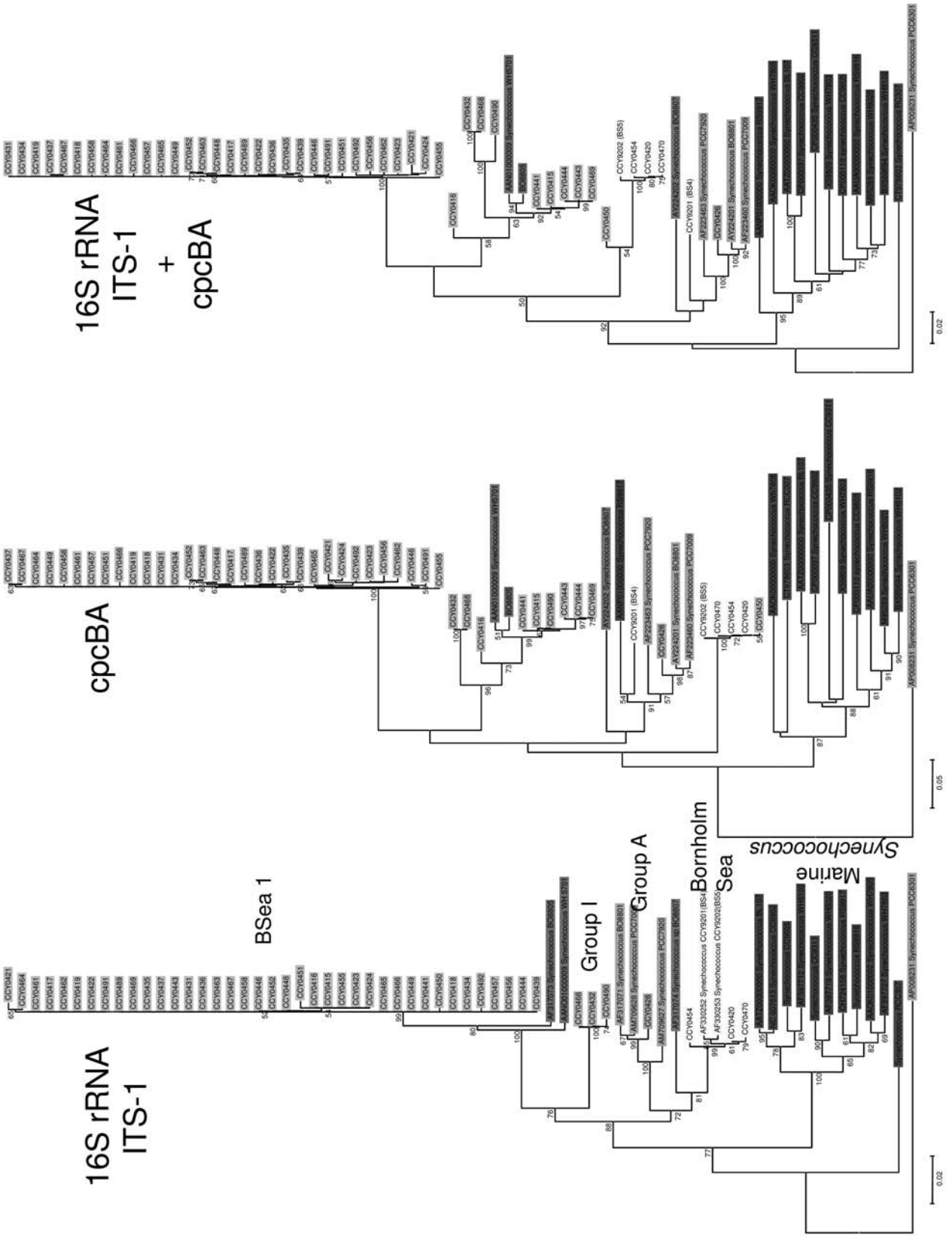

16S rRNA
ITS-1
+
cpcBA
cpcBA
16S rRNA
ITS-1
BSea 1
Group I
Group A
Bornholm
Sea
Marine
Synechococcus
0.02
0.05
0.02

Figure SM6

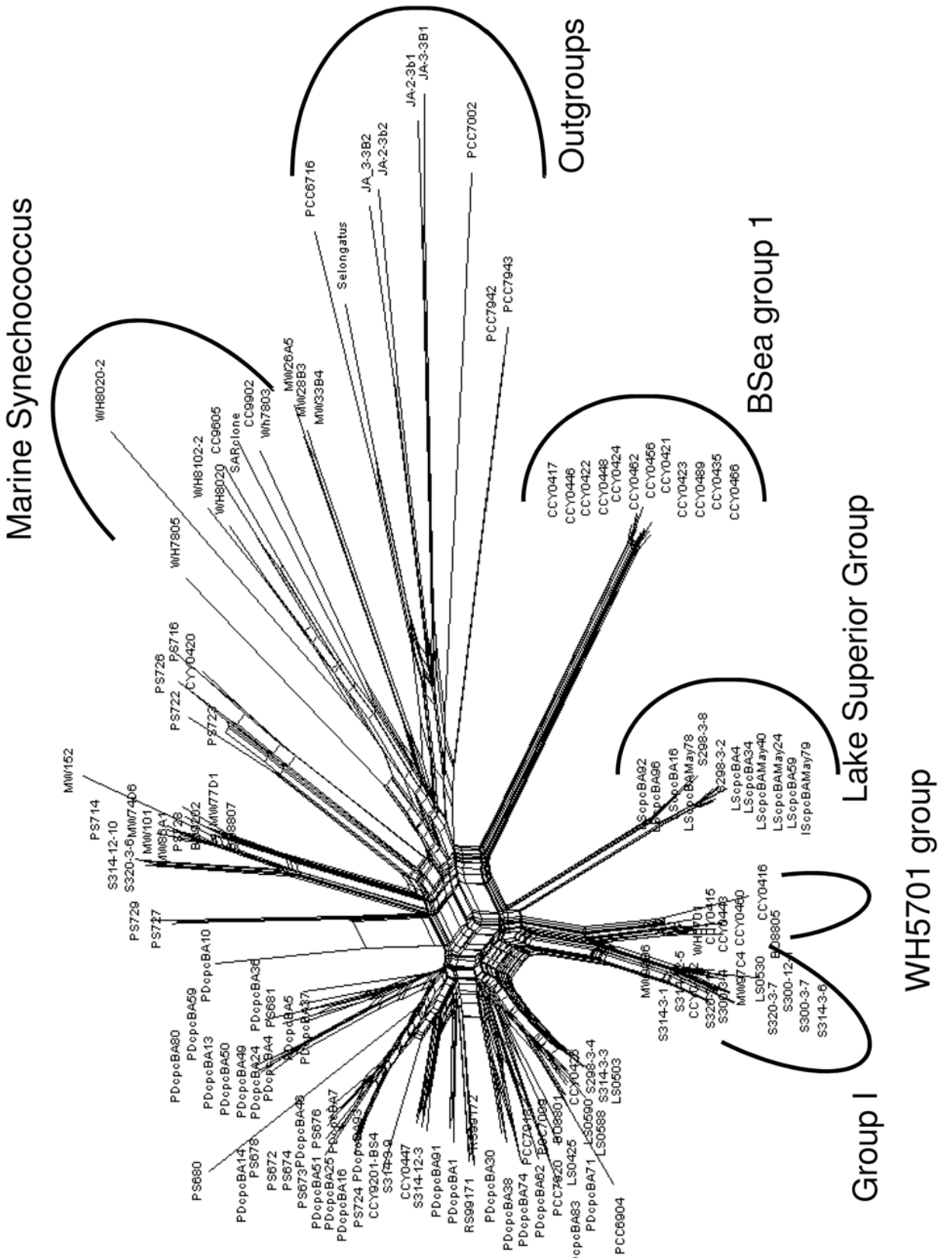

Marine Synechococcus
Outgroups
BSea group 1
Lake Superior Group
WH5701 group
Group I
WH8020-2
PCC6716
Selongatus
JA_3-3B2
JA-2-3b2
JA-2-3b1
JA-3-3B1
PCC7002
PCC7942
PCC7943
MW26A5
MW28B3
MW33B4
WH7805
WH8102-2
WH8020
CC9605
SARclone
CC9902
Wh7803
PS726
PS716
CCY0420
PS722
PS723
MW152
PS714
S314-12-10
S320-3-6
MW7406
MW101
MW077D1
S807
PS729
PS727
PDcpcBA80
PDcpcBA59
PDcpcBA10
PDcpcBA13
PDcpcBA50
PDcpcBA49
PDcpcBA24
PDcpcBA4
PDcpcBA36
PS681
PDcpcBA5
PDcpcBA37
PS680
PDcpcBA14
PS678
PS672
PS674
PS673
PDcpcBA46
PS676
PDcpcBA51
PDcpcBA25
PDcpcBA16
PS724
CCY9201-BS4
S314-3-9
CCY0447
S314-12-3
PDcpcBA91
PDcpcBA1
RS99171
PDcpcBA30
PDcpcBA38
PDcpcBA74
PDcpcBA62
PCC7920
PDcpcBA83
LS0425
PDcpcBA71
PCC6904
LS0590
LS0588
S298-3-4
S314-3-3
LS0503
CCY0426
CCY0417
CCY0446
CCY0422
CCY0448
CCY0424
CCY0462
CCY0456
CCY0421
CCY0423
CCY0489
CCY0435
CCY0466
SCpcBA92
SCpcBA96
SCpcBA16
LSCpcBAMay78
S298-3-8
S298-3-2
LScpcBA4
LScpcBA34
LScpcBAMay40
LScpcBAMay24
LScpcBA59
lScpcBAMay79
CCY0416
CCY0448
CCY0440
LS0530
S320-3-7
S300-12-
S300-3-7
S314-3-6

Figure SM7