The ISME Journal (2009) 3, 397–408

www.nature.com/ismej

ORIGINAL ARTICLE

Colorful microdiversity of *Synechococcus* strains (picocyanobacteria) isolated from the Baltic Sea

Thomas HA Haverkamp[1], Daphne Schouten[1], Marije Doeleman[1], Ute Wollenzien[1], Jef Huisman[2] and Lucas J Stal[1,2]

[1]*Department of Marine Microbiology, Netherlands Institute of Ecology (NIOO-KNAW), Yerseke, The Netherlands and* [2]*Department of Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands*

Synechococcus **is a cosmopolitan genus of picocyanobacteria living in the photic zone of freshwater and marine ecosystems. Here, we describe the isolation of 46 closely related picocyanobacterial strains from the Baltic Sea. The isolates showed considerable variation in their cell size and pigmentation phenotypes, yielding a colorful variety of red, pink and blue-green strains. These pigmentation phenotypes could not be differentiated on the basis of their 16S rRNA-internal transcribed spacer (ITS) sequences. Thirty-nine strains, designated BSea, possessed 16S rRNA-ITS sequences almost identical with** ***Synechococcus*** **strain WH5701. Despite their similar 16S rRNA-ITS sequences, the BSea strains separated into several different clusters when comparing the phycocyanin (*cpcBA*) operon. This separation was largely consistent with the phycobiliprotein composition of the different BSea strains. The majority of phycocyanin (PC)-rich Bsea strains clustered with WH5701. Remarkably, the phycoerythrin (PE)-rich strains of BSea 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 indicated that the PE-rich BSea strains are probably endemic for the Baltic Sea. Comparison of the phylogenies obtained by the 16S rRNA-ITS, the *cpcBA*, and 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 extensive diversification of different pigmentation phenotypes into different ecological niches.**

The ISME Journal (2009) **3,** 397–408; doi:10.1038/ismej.2008.118; published online 4 December 2008
Subject Category: microbial population and community ecology
Keywords: horizontal gene transfer; microdiversity; phycocyanin; *Synechococcus*; cyanobacteria; phycoerythrin

Introduction

Synechococcus is a typical cosmopolitan genus comprising both picocyanobacteria (cells $< 2\,\mu m$) as well as larger cyanobacteria (Schmidt *et al.*, 1991). The genus is polyphyletic and genetically highly diverse. *Synechococcus* species are phylogenetically divided into several major clusters. Picocyanobacteria that are often found and isolated from marine, brackish and freshwater environments are related to *Synechococcus* cluster 5 (Herdman *et al.*, 2001; 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*.

Correspondence: LJ Stal, Department of Marine Microbiology, Netherlands Institute of Ecology (NIOO-KNAW), PO Box 140, Yerseke 4400 AC, The Netherlands.
E-mail: L.Stal@nioo.knaw.nl
Received 31 October 2008; accepted 1 November 2008; published online 4 December 2008

The diversity of *Synechococcus* and its sister groups has been investigated mainly by analysis of the 16S rRNA gene obtained from environmental clone libraries or from strains isolated from various environments throughout the world. However, the 16S rRNA gene is conserved and has limited resolution. Several studies have therefore used less conserved genetic markers, such as the internal transcribed spacer (ITS) separating the 16S and 23S ribosomal genes or the phycocyanin operon (*cpcBA*), to study the diversity of picocyanobacteria (for example, Rocap *et al.*, 2002; Crosbie *et al.*, 2003; Everroad and Wood, 2006). In particular, isolated strains offer the possibility to concatenate multiple genetic markers, which can provide a detailed picture of the evolutionary history and genetic diversity of picocyanobacteria.

Synechococcus consists of strains rich in the pigment phycoerythrin (PE), rendering its representatives a variety of orange, reddish, pink and purple

colors, and strains rich in phycocyanin (PC), coloring the organism in various shades of blue-green. Competition experiments showed that PE-rich and PC-rich strains can coexist in white light, through niche differentiation in the light spectrum (Stomp *et al.*, 2004). This spectral niche differentiation can also explain the distribution of reddish and blue-green strains in natural waters. PE-rich strains dominate in clear waters; they contain relatively high contents of the chromophore phycourobilin (PUB) in the clearest ocean waters in which blue light prevails, whereas their chromophore composition shifts towards phycoerythrobilin (PEB) in more mesotrophic marine waters characterized by blue-green light environments (Olson *et al.*, 1990; Lantoine and Neveux, 1997; Wood *et al.*, 1998; Scanlan, 2003). Conversely, PC-rich strains dominate in turbid inland waters in which orange and red light prevail (Pick, 1991; Vörös *et al.*, 1998; Stomp *et al.*, 2007a, b). Widespread coexistence of PE- and PC-rich picocyanobacteria can be found in waters of intermediate turbidity, such as mesotrophic lakes and coastal seas (Stomp *et al.*, 2007a; Haverkamp *et al.*, 2008). On the basis of these previous studies, we hypothesize that waters of intermediate turbidity, such as the Baltic Sea, are hotspots for the pigmentation diversification of picocyanobacteria.

In this study, the partial 16S rRNA, the ITS region and the *cpcBA* operon sequences were concatenated to study the phylogenetic relationship among a colorful diversity of *Synechococcus* strains isolated from the Baltic Sea and other known picocyanobacteria. The data sets were analyzed phylogenetically and by rarefaction. We show that the concatenated data set yielded higher estimates of cyanobacterial diversity than the analyses of the data sets separately. Furthermore, the data suggest geographical endemism and horizontal gene transfer, indicative of rapid diversification of *Synechococcus* in the Baltic Sea.

Materials and methods

Isolation of strains

Strains were isolated from water samples collected during a research cruise with the Finnish RV Aranda in the Gulf of Finland (Baltic Sea) from 12 to 19 July 2004. Details on the sampling sites and isolation procedure are given in the Supplementary Information(Supplementary Figures SM1 and SM2; Supplementary Tables SM1 and SM2). Isolated strains were maintained in liquid BSea6 medium, which is a mixture of one part ASNIII and four parts BG11 medium (Rippka *et al.*, 1979) with a final salinity of 6.0 psu representative of the brackish waters of the Baltic Sea. We modified our BSea6 medium to obtain a final concentration of 1 mM NH_4Cl, 1.8 mM $NaNO_3$ and 0.012 mM $NaHCO_3$.

Morphology and microscopy

Microscope slides were prepared by covering them with a thin layer of 1% hot agarose (50 °C) that was subsequently allowed to solidify. Cells of the isolated strains were applied to the agarose and covered with a coverslip. Observations were performed using a Zeiss Axiophot microscope (Zeiss, 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). At least 40 measurements of cell width (W) and length (L) were taken and averaged for each culture. Cell volume (V) 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 (4 ml) and the *in vivo* absorption spectrum was measured from 400 to 750 nm using a Varian Cary 100 Bio equipped with an integrating sphere DRA-CA-3300 (Varian, Palo Alto, CA, USA).

DNA extraction

Stationary phase cultures were used for DNA extraction. One ml of culture was transferred to a microcentrifuge tube provided with 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 (10 mM Tris, 1 mM EDTA, pH 8.0). The DNA concentration was measured using an 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), whereas 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), 0.2 mM of each dNTP (Roche, Woerden, The Netherlands), 1 × PCR buffer (Qiagen), 5 pmol of the forward and reverse primers were used per reaction and purified Milli-Q-Grade (MQ-grade) water (Millipore purification unit, 18 MΩ · cm (Millipore, Billerica, MA, USA)) 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 using electrophoresis with a TAE buffer (pH 8.5) on a 1% agarose (Sigma Aldrich, Zwijndrecht, The Netherlands) gel.

Cloning and sequencing

PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen, Breda, The Netherlands) following the instructions of the manufacturer. Five colonies were picked for each PCR product and inoculated in microtiterplates (type 655180, Greiner Bio-one, Alphen a/d Rijn, The Netherlands) 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 to amplify the inserted gene fragments. The PCR reaction mixture contained 0.625 units HotStarTaq (Qiagen), 0.2 mM of each DNTP (Roche), 1 × 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 s at 94 °C, 30 s 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 and 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 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). In case of nucleotide differences, chromatograms were checked for base calling errors, and the sequences were corrected. The clone sequences were then aligned in BioEdit and a consensus sequence was created for each strain, using ambiguous bases for the positions with nucleotide differences. 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 nos. EU386608–EU386699.

Phylogenetic and diversity analysis

Phylogenetic analysis was performed using alignments of the 16S rRNA-ITS-1 and the *cpcBA* operon sequences of our isolated strains and 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 was removed from the *cpcBA* operon sequences before analysis for reasons of reliable alignment. 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 (ML) analysis using the 16S rRNA-ITS-1 region and the *cpcBA* operon was performed using PAUP* (Swofford, 2003). The ML of the 16S rRNA-ITS tree was evaluated using the Jukes–Cantor model of nucleotide substitution, whereas the ML of the *cpcBA* tree was evaluated using the Kimura two parameter model. The Kimura model assumes that transition and transversion mutations may occur at different rates, which is an appropriate assumption for protein-coding sequences such as the *cpcBA* operon. If transition rates equal transversion rates, which is a reasonable assumption for non-coding sequences such as the 16S rRNA-ITS, the Kimura model simplifies to the Jukes–Cantor model (Huelsenbeck, 1995). The congruence of the 16S rRNA-ITS and the *cpcBA* tree was investigated by fitting the sequences of one gene to the ML tree of the other gene, and vice versa, followed by re-sampling (Swofford, 2003). The resulting distributions of log-likelihoods and associated *P*-values were compared to assess whether the two phylogenetic trees were significantly different.

Analysis of the diversity at the different loci among the picocyanobacterial isolates was performed using rarefaction analysis with the software DOTUR (Schloss and Handelsman, 2005). As input for rarefaction analysis a distance matrix of the sequence alignments of the different loci was created 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

Morphology and pigmentation of the isolates

In total, we isolated 46 picocyanobacteria from the Baltic Sea. The strains, their Culture Collection Yerseke (CCY) number, clade assignment, station and depth of isolation, pigmentation, size and volumes are listed in Table 1. Cell dimensions and biovolumes varied considerably between the isolates. The coccoid isolate CCY0452 was the smallest among the isolates (biovolume of 0.66 μm^3) and the rod CCY0454 was the largest (biovolume of 2.43 μm^3). Cell width varied only little among strains (0.8–1.4 μm), whereas cell length varied from 1.2–2.9 μm causing the biovolume differences. The isolates differed widely in color, including various shades of red, pink and blue-green strains. Thirteen strains contained large amounts of PC but lacked PE (Table 1). The high PC content of these PC-rich strains is illustrated by their light absorption spectra, as major peaks at 620 nm (Figure 1). The other 33 strains contained large amounts of PE but little PC. The predominant chromophore of the PE-rich strains was PEB, yielding high absorption peaks at 565 nm. In addition, some of the PE-rich strains also contained small amounts of the PUB, indicated by minor absorption peaks at 495 nm. The absorption spectra of the 46 isolated strains showed considerable phenotypic variation with respect to the ratio of phycobiliprotein (either PE or PC) to Chl *a*, despite the fact that they were maintained under the same growth conditions.

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 16S rRNA-ITS-1 region, all strains were highly similar to known picocyanobacteria (Table 1 and Figure 2). The isolates belong to four 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 picocyanobacteria (Ernst *et al.*, 2003; Chen *et al.*, 2006). Group A is also known as the *Cyanobium gracile* cluster (Ernst *et al.*, 2003). The second group comprises the strains CCY0420, CCY0454 and CCY0470. Their 16S rRNA-ITS-1 sequences form a monophyletic clade with CCY9201 (former BS4) and CCY9202 (former BS5) that were previously isolated from the Baltic Sea (Ernst *et al.*, 2003; Stomp *et al.*, 2004) and were assigned as 'Bornholm Sea Group'. The third clade comprises the ITS-1 sequences of the strains CCY0432, CCY0468 and CCY0490. These three strains form another deep branching phylogenetic cluster related with the 16S rRNA sequences of Group I (Crosbie *et al.*, 2003). The fourth and largest group contains 39 strains and is designated the Baltic Sea group (Bsea; Figure 2). 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 subalpine cluster II (Ernst *et al.*, 2003; Chen *et al.*, 2006). The 16S rRNA-ITS sequences of BSea isolates can be completely aligned with WH5701 and BO8805. Nonetheless, the sequences of the BSea 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 BSea 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 BSea group, microdiversity is observed when the full-length ITS-1 sequences are used for phylogenetic analysis (Supplementary Figure SM3). Four strains form a distinct cluster (BSeaB) that has on average 1.3% sequence difference with the other major subcluster, BSeaA (Supplementary Figure SM3). Originally, *Synechococcus* WH5701 was assigned to the marine *Synechococcus* cluster 5.2 (Herdman *et al.*, 2001). However, based on the similarity of the 16S rRNA and the ITS sequences with strain BO8805, WH5701 can also be assigned to the subalpine cluster II (Crosbie *et al.*, 2003; Ernst *et al.*, 2003; Chen *et al.*, 2006). Interestingly, in the taxonomic description of marine *Synechococcus* cluster 5.2, all members of this clade were assumed to have PC as their main light-harvesting pigment (that is, they are blue-green) (Herdman *et al.*, 2001). However, as shown here, strains belonging to subalpine cluster II may also possess PE. In fact, many of our Baltic Sea isolates closely related to BO8805 and WH5701 contain PE as their main light-harvesting pigment. This shows that pigmentation does not necessarily correspond to the 16S rRNA and ITS-1 sequences.

PC diversity of picocyanobacteria

Earlier work has shown that the *cpcBA* operon of picocyanobacteria can separate strains into different phylogenetic groups with clearly different pigment phenotypes (Six *et al.*, 2007; Haverkamp *et al.*, 2008). Indeed, comparison of the *cpcBA* operon sequences of our 46 isolates and those available in GenBank shows that the majority of the Baltic Sea isolates clustered according to pigment phenotype (Supplementary Figure SM4). This is especially clear for the strains belonging to the BSea group, which separated into 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 when based on the phylogeny of the *cpcBA* operon (Supplementary Figure SM4) (Crosbie *et al.*, 2003; Haverkamp *et al.*, 2008). Nonetheless, a more detailed representation of this phylogenetic tree shows that there are exceptions. Some PC-rich strains cluster with PE-rich isolates

Table 1 Picocyanobacteria isolated in this study

Strain	*Clade*	*Site of isolation*	*Station*	*Isolation depth (m)*	*Cell length (μm)*	*Cell width (μm)*	*Cell biovolume ($μm^3$)*	*Major phycobiliprotein*
CCY0415	Subalpine C II	59° 30′ N, 22° 40′ E	S314	0	1.8 ± 0.5	1.2 ± 0.2	1.5 ± 0.6	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 ± 0.3	1.1 ± 0.1	1.1 ± 0.5	PC
CCY0418	Subalpine C II	59° 30′ N, 22° 40′ E	S314	0	2.1 ± 0.6	1.1 ± 0.1	1.8 ± 0.6	PE
CCY0419	Subalpine C II	59° 30′ N, 22° 40′ E	S314	0	1.6 ± 0.4	1.1 ± 0.1	1.2 ± 0.4	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 ± 0.2	1.1 ± 0.3	0.9 ± 0.6	PE
CCY0422	Subalpine C II	59° 30′ N, 22° 40′ E	S314	15	1.6 ± 0.3	1.1 ± 0.1	1.1 ± 0.3	PE
CCY0423	Subalpine C II	59° 13′ N, 22° 19′ E	S327	15	1.9 ± 0.6	1.1 ± 0.1	1.5 ± 0.9	PE
CCY0424	Subalpine C II	59° 13′ N, 22° 19′ E	S327	15	1.6 ± 0.3	1.1 ± 0.1	1.1 ± 0.4	PE
CCY0426	Group A *Cyanobium*	59° 30′ N, 22° 43′ E	S310	0	2.0 ± 0.4	1.3 ± 0.1	1.9 ± 0.6	PC
CCY0431	Subalpine C II	59° 30′ N, 22° 40′ E	S314	15	1.7 ± 0.5	1.1 ± 0.1	1.3 ± 0.7	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 ± 0.4	1.2 ± 0.2	1.2 ± 0.7	PE
CCY0435	Subalpine C II	59° 30′ N, 22° 40′ E	S314	0	1.8 ± 0.5	1.1 ± 0.2	1.3 ± 0.5	PE
CCY0436	Subalpine C II	59° 24′ N, 22° 26′ E	S322	0	1.7 ± 0.5	1.1 ± 0.2	1.3 ± 0.6	PE
CCY0437	Subalpine C II	59° 13′ N, 22° 19′ E	S327	0	1.7 ± 0.4	1.2 ± 0.1	1.5 ± 0.5	PC
CCY0439	Subalpine C II	59° 30′ N, 22° 40′ E	S314	0	1.6 ± 0.4	1.1 ± 0.1	1.2 ± 0.5	PE
CCY0441	Subalpine C II	59° 28′ N, 22° 39′ E	S305	0	1.8 ± 0.5	1.1 ± 0.1	1.3 ± 0.5	PC
CCY0443	Subalpine C II	59° 28′ N, 22° 39′ E	S305	0	2.0 ± 0.4	1.0 ± 0.1	1.2 ± 0.4	PC
CCY0444	Subalpine C II	59° 30′ N, 22° 50′ E	S307	0	2.1 ± 0.5	1.1 ± 0.1	1.6 ± 0.5	PE
CCY0446	Subalpine C II	59° 30′ N, 22° 43′ E	S310	0	1.3 ± 0.3	1.1 ± 0.1	0.9 ± 0.3	PE
CCY0448	Subalpine C II	59° 30′ N, 22° 43′ E	S310	12	1.5 ± 0.2	1.4 ± 0.2	1.6 ± 0.5	PE
CCY0449	Subalpine C II	59° 30′ N, 22° 43′ E	S310	12	2.0 ± 0.4	1.1 ± 0.1	1.6 ± 0.4	PE
CCY0450	Subalpine C II	59° 30′ N, 22° 43′ E	S310	12	1.6 ± 0.3	1.0 ± 0.1	1.0 ± 0.3	PE
CCY0451	Subalpine C II	59° 30′ N, 22° 43′ E	S310	12	1.7 ± 0.3	1.0 ± 0.1	1.1 ± 0.4	PE
CCY0452	Subalpine C II	59° 30′ N, 22° 43′ E	S310	12	1.2 ± 0.3	1.0 ± 0.1	0.7 ± 0.2	PC
CCY0454	Bornholm Sea	59° 30′ N, 22° 43′ E	S310	0	2.9 ± 0.7	1.1 ± 0.2	2.5 ± 1.0	PE
CCY0455	Subalpine C II	59° 30′ N, 22° 40′ E	S314	0	2.1 ± 0.4	1.1 ± 0.1	1.6 ± 0.4	PC
CCY0456	Subalpine C II	59° 32′ N, 22° 50′ E	S301	0	1.5 ± 0.3	1.1 ± 0.1	1.1 ± 0.4	PE
CCY0457	Subalpine C II	59° 30′ N, 22° 40′ E	S314	0	1.7 ± 0.3	1.0 ± 0.1	1.1 ± 0.3	PE
CCY0458	Subalpine C II	59° 30′ N, 22° 40′ E	S314	0	1.9 ± 0.4	1.0 ± 0.1	1.1 ± 0.3	PE
CCY0461	Subalpine C II	59° 32′ N, 22° 50′ E	S301	0	1.5 ± 0.3	0.9 ± 0.1	0.8 ± 0.3	PE
CCY0462	Subalpine C II	59° 30′ N, 22° 40′ E	S314	15	1.4 ± 0.2	1.0 ± 0.1	0.8 ± 0.2	PC
CCY0463	Subalpine C II	59° 30′ N, 22° 40′ E	S314	15	1.7 ± 0.3	1.1 ± 0.1	1.2 ± 0.4	PE
CCY0464	Subalpine C II	59° 30′ N, 22° 40′ E	S314	15	1.6 ± 0.4	1.2 ± 0.1	1.3 ± 0.4	PE
CCY0465	Subalpine C II	59° 30′ N, 22° 40′ E	S314	15	1.8 ± 0.4	1.1 ± 0.1	1.3 ± 0.4	PE
CCY0466	Subalpine C II	59° 24′ N, 22° 26′ E	S322	0	1.3 ± 0.1	1.1 ± 0.1	0.9 ± 0.3	PE
CCY0467	Subalpine C II	59° 24′ N, 22° 26′ E	S322	0	1.4 ± 0.2	1.2 ± 0.1	1.1 ± 0.4	PE
CCY0468	Group I	59° 32′ N, 22° 50′ E	S301	0	1.8 ± 0.4	0.8 ± 0.1	0.8 ± 0.3	PE
CCY0469	Subalpine C II	59° 24′ N, 22° 26′ E	S322	15	1.3 ± 0.2	1.0 ± 0.1	0.8 ± 0.3	PC
CCY0470	Bornholm Sea	59° 28′ N, 22° 39′ E	S305	0	1.7 ± 0.4	1.1 ± 0.1	1.2 ± 0.5	PE
CCY0489	Subalpine C II	59° 28′ N, 22° 39′ E	S305	0	1.7 ± 0.6	1.0 ± 0.2	1.2 ± 0.5	PE
CCY0490	Group I	59° 28′ N, 22° 39′ E	S305	0	1.8 ± 0.3	0.9 ± 0.1	0.9 ± 0.2	PC
CCY0491	Subalpine C II	59° 28′ N, 22° 39′ E	S305	0	1.7 ± 0.5	1.0 ± 0.1	1.1 ± 0.5	PE
CCY0492	Subalpine C II	59° 28′ N, 22° 39′ E	S305	0	1.3 ± 0.3	1.0 ± 0.1	0.8 ± 0.3	PC

Abbreviations: ITS, internal transcribed spacer; ND, not determined; PC, phycocyanin; PE, phycoerythrin; PEB, phycoerythrobilin; PUB, phycourobilin.
Clade designation is based on the highest similarity between the 16S rRNA-ITS sequence of the isolate and known sequences found in GenBank. Cell length and width (± standard deviation) were measured on 40 cells. Biovolumes were calculated from the length and width of each cell using the formula for a cylinder with hemispherical ends (Sieracki *et al.*, 1989). Strains indicated by PC contained large amounts of phycocyanin but no phycoerythrin, whereas strains indicated by PE contained large amounts of phycoerythrin and only a little phycocyanin. All PE-rich strains contained PEB as predominant chromophore, and some of these strains also contained a little PUB.

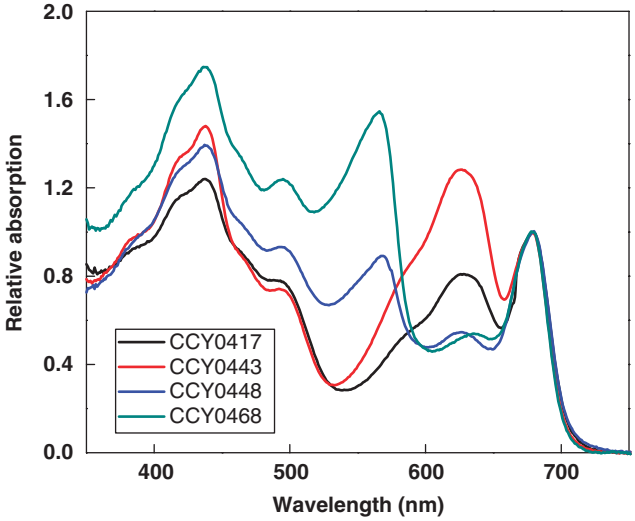

Figure 1 Comparison of *in vivo* absorption spectra of two phycoerythrin-rich and two phycocyanin-rich Baltic Sea isolates grown under the same light conditions. For comparison, the spectra were normalized at the red peak of chlorophyll *a* (680 nm). Despite striking differences in phycobilisome composition, the strains CCY0417 and CCY0443 belong to the same phylogenetic clade according to their 16S rRNA-ITS-1 sequences.

of the BSea group and vice versa (Supplementary Figure SM5). Remarkably, the PE-rich isolates CCY0432 and CCY0468 fall into the Baltic Sea cluster of environmental clones in Group I, which until now contained only PC-rich isolates. The *cpcBA* sequences of PE-rich strains CCY0420, CCY0450, CC0454 and CCY0470 form a monophyletic clade with the strains belonging to the PE-rich group E including the Bornholm Sea strain CCY9202. Finally, the *cpcBA* operon sequence of PC-rich strain CCY0426 clusters with the PC-rich group A (Crosbie *et al.*, 2003) (Supplementary Figure SM5).

Comparison of the *cpcBA* data from the environmental clone libraries extracted in our previous study (Haverkamp *et al.*, 2008) and from the strains isolated in the presented study revealed a much higher diversity in the combined dataset than in either the clone libraries or the isolates alone (Table 2; Supplementary Figure SM5). Phylogenetic analysis revealed that this is caused by the presence of several sequences that were found in only one of the two data sets, thus increasing the overall diversity (Supplementary 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, the available data of the *cpcBA* operon distinguished 8 operational taxonomic units (OTUs) in the Baltic Sea, and 31 OTUs worldwide when using a 90% similarity cutoff (Table 2).

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

The high resolution of the *cpcBA* phylogeny makes it interesting to compare the phylogenetic trees derived from the 16S rRNA-ITS sequences (Figure 2) and *cpcBA* sequences (Supplementary Figures SM4 and SM5). However, a direct comparison between the two phylogenetic trees is difficult for the following reasons. (1) The number of sequences in the two phylogenetic trees is unequal. (2) Some strains are present in only one of the two trees, because their sequence data for the other operon are lacking. (3) In order to construct the 16S rRNA-ITS-1 phylogenetic tree, large variable regions had to be removed. Therefore, in addition to the sequence data from our own isolated strains, we selected only those strains from GenBank for which both the *cpcBA* sequences and 16S rRNA-ITS-1 sequences were available. 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, however there were several exceptions (Figure 3). This is especially clear for strain CCY0450, which belongs to the BSea group with respect to the 16S rRNA-ITS-1 sequence, 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 BSea. Other strains that behave in a similar way are CCY0416 (BSea), CCY0490 (Group I), CCY9201 (Bornholm Sea) and the marine *Synechococcus* strains RS9917 and RCC307 (Figure 3) (Fuller *et al.*, 2003; Six *et al.*, 2007; Haverkamp *et al.*, 2008).

Interestingly, in the *cpcBA* phylogeny the BSea group is split into two groups. Several of the PC-rich strains of the BSea group (CCY0415, CCY0416, CCY0441, CCY0443, CCY0469) and one PE-rich strain (CCY0444) cluster within the *cpcBA* phylogeny with the PC-rich strains CCY0490, WH5701 and BO8805 (Figure 3). This contradicts the 16S rRNA-ITS-1 phylogeny in which PE-rich strains of the BSea group form a monophyletic clade with PC-rich strains including WH5701 and BO8805. Hence, this visual comparison of the 16S rRNA-ITS and *cpcBA* trees indicates a lack of congruency between the two phylogenies. Despite our effort to equalize the number of sequences between the two phylogenies, it was still difficult to compare the phylogenetic trees of 16S rRNA-ITS and *cpcBA* by visual inspection only. Therefore, we performed a detailed statistical analysis for those clades for which we can compare the full length ITS sequences with each other. These clades are the BSea group, WH5701, BO8805 and the Group I strains. For these clades we can test if the topologies of the two trees are different using ML. For this purpose, the likelihoods of the 16S rRNA-ITS and the *cpcBA* trees were both estimated using the Jukes–Cantor model of

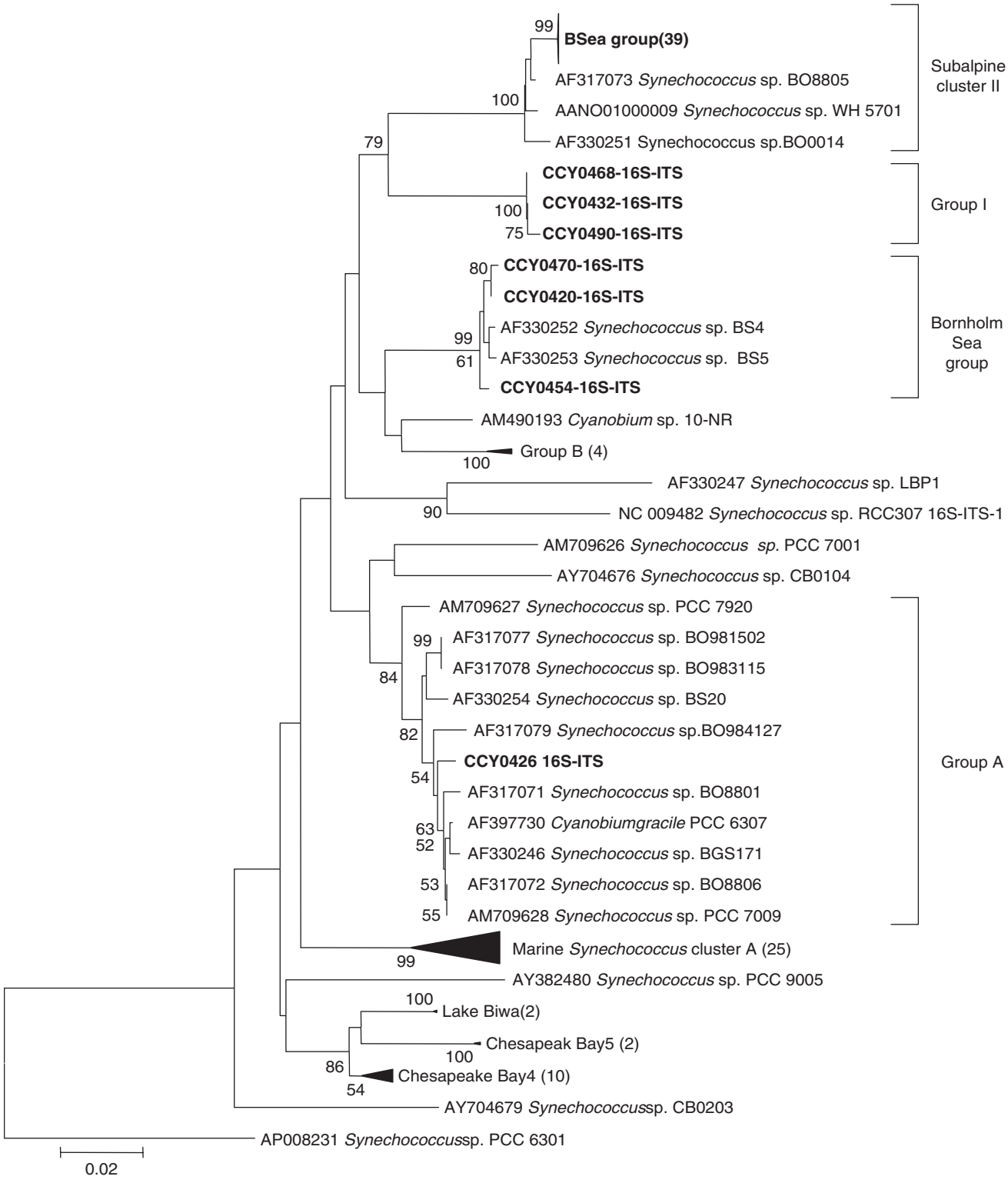

Figure 2 Neighbour-joining tree based on the 16S rRNA-ITS-1 sequences from 112 picocyanobacterial isolates. The partial sequences contained the highly conserved regions of the 16S rRNA and the ITS-1 region (862 nucleotides). The tree was rooted with the sequence of *Synechococcus* PCC6301. Construction of the neighbour-joining tree was based on maximum likelihood using the Jukes–Cantor model of nucleotide substitution. Bootstrap values (1000 replicates) are shown for confidence levels higher than 50%. Picocyanobacterial isolates from the Baltic Sea are identified using bold characters. For clustered taxa, the number of strains per cluster is indicated between parentheses.

nucleotide substitution. Subsequently, the congruence of the two trees was evaluated by fitting the 16S rRNA-ITS sequences to the ML tree of the *cpcBA* operon, and vice versa, followed by re-sampling (Swofford, 2003). Comparison of the resulting distributions of log-likelihoods showed that the 16S rRNA-ITS sequences fitted significantly better to the 16S rRNA-ITS tree than to the *cpcBA* tree ($P<0.001$). Conversely, the *cpcBA* sequences fitted significantly better to the *cpcBA* tree than to the 16S rRNA-ITS tree ($P<0.001$). This showed that the two phylogenetic trees were indeed not congruent.

Table 2 Comparison of diversity of picocyanobacterial *cpcBA* operon obtained from different data sets. Number of operational taxonomic units (OTUs) is given at the similarity cutoff levels 100, 99, 97, 94 and 90%

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

[a]Data from Haverkamp *et al.* (2008).

Finally, both data sets were used to create concatenated sequences of the 16S rRNA-ITS-1 and the *cpcBA* operon. Most clades found in the *cpcBA* phylogeny appeared also in the concatenated data set (Figure 3). Yet, in the concatenated data set CCY0490 and RS9917 cluster with strains with which they share 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 with the *cpcBA* alone (Figure 3). This conclusion is

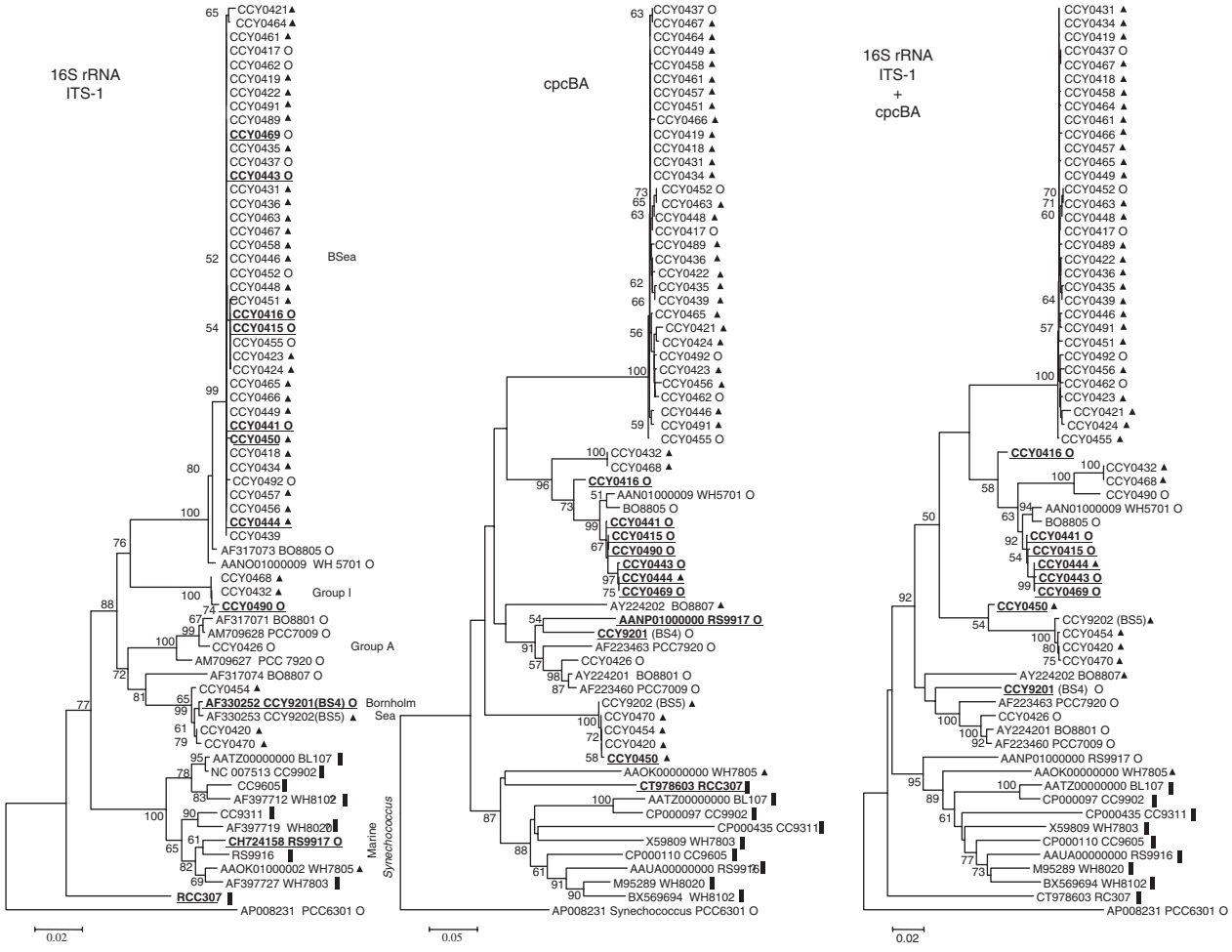

Figure 3 Comparison of the phylogenetic trees of the 16S rRNA-internal transcribed spacer (ITS) region, the *cpcBA* operon and the concatenated sequences. The symbols behind the strain numbers represent the pigmentation of the strains, including phycocyanin-rich strains (○), phycoerythrin (PE)-rich strains with low phycourobilin (PUB)/phycoerythrobilin (PEB) ratios (▲), and PE-rich strains with high PUB/PEB ratios (■). 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. The *cpcBA* phylogeny was obtained using an alignment of the *cpcBA* operon without the intergenic spacer (IGS) (498 nucleotides). The trees were constructed with a neighbour-joining algorithm, using the Jukes–Cantor model of nucleotide substitution for the non-functional 16S rRNA-ITS sequences, and the Kimura two parameter model for the protein-coding *cpcBA* operon and the concatenated sequences. The resulting trees were bootstrapped 1000 times for statistical support.

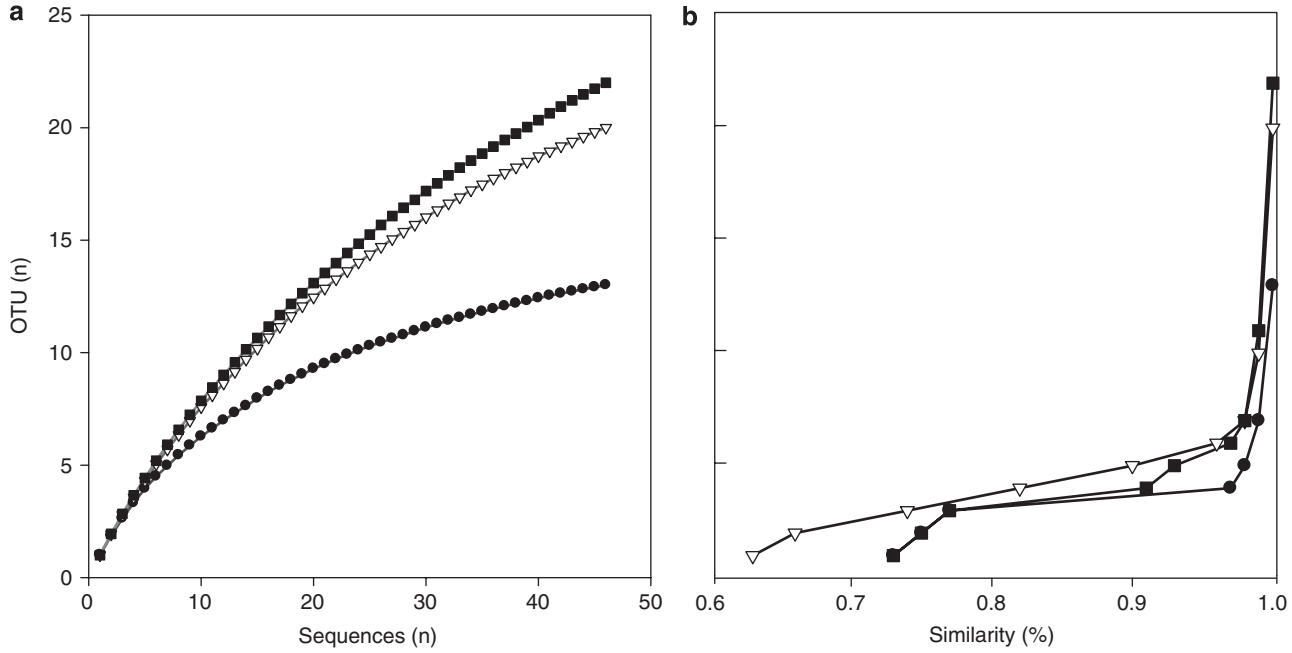

Figure 4 Comparison of diversity at different loci among the Baltic Sea isolates. (**a**) 100% similarity rarefaction curves, and (**b**) similarity plots of the 16S rRNA-ITS sequences (closed circles), *cpcBA*-operon (open triangles) and the concatenated 16S rRNA-ITS and *cpcBA* operon sequences (closed squares).

supported by rarefaction analysis of the data sets, where concatenation increased the diversity at the 100% similarity level to 22 OTUs compared with 20 OTUs for the *cpcBA* operon alone and 13 OTUs for the 16S rRNA-ITS-1 alone, respectively (Figure 4).

Split network analysis

The *cpcBA* phylogeny is represented by well-supported clades deep in the tree (Supplementary Figures SM4 and SM5). However, the basal taxa of the *cpcBA* phylogeny are shown with long branches and the nodes have low bootstrap support (<50%) (Supplementary 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 tree. One way to investigate this is by analyzing the *cpcBA* data set 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 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 5; the same analysis using the complete set of sequences can be found in Supplementary Figure SM6). 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. 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, by the long branches the neighbour-net network shows that the *cpcBA* sequences of the PE-rich BSea cluster (represented by strain CCY0418) and the other groups form distinct taxa within the *cpcBA* phylogeny of the picocyanobacteria (Figure 5 and Supplementary Figure SM6). In addition, it shows a clear separation between the *cpcBA* genes of PE-rich and PC-rich picocyanobacteria (Figure 5).

Discussion

Isolation of 46 *Synechococcus* strains from the Baltic Sea provided detailed information on the microdiversity of these picocyanobacteria. The strains were compared using morphological and molecular approaches. At the morphological level the most striking difference between isolates was their color. The strains were rich in either PE or PC. The PE-rich strains could not be differentiated from the PC-rich strains on the basis of their 16S rRNA-ITS sequences. However, this clearly visible phenotypic differentiation was genetically supported by the *cpcBA* operon, which separated many of the PE-rich strains from the PC-rich strains (Figure 3; see also Six *et al*., 2007; Haverkamp *et al*., 2008). Other phenotypic differences, in cell size and cell volume, did not match the separation into different phylogenetic clusters.

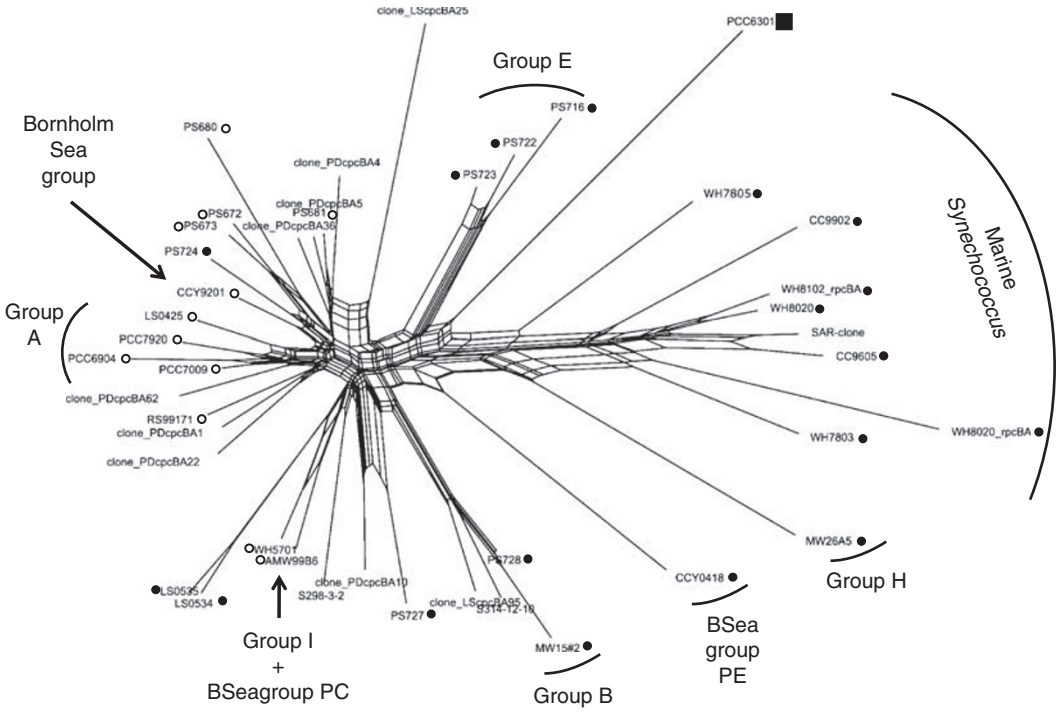

Figure 5 A neighbour-network constructed using 42 representative *cpcBA* sequences (determined by rarefaction analysis with a cutoff of 94%). Bootstrap support values for the nodes are omitted for clarity. The out-group is represented by the *cpcBA* sequence of *Synechococcus* PCC6301 (■). The phycoerythrin (PE)-rich BSea1 cluster is represented by strain CCY0418. PE-rich strains are indicated by closed circles (●), whereas PC-rich strains are indicated by open circles (○). For a complete analysis using all known *cpcBA* sequences see Supplementary Figure SM5.

The colorful diversity of Baltic Sea Synechococcus

Our results confirm that coexistence of PE-rich and PC-rich strains is widespread in the Baltic Sea (Stomp *et al.*, 2004, 2007a; Haverkamp *et al.*, 2008). In addition, the isolates presented in this study show that PE-rich strains in the Baltic Sea contain PEB as major chromophore, whereas PUB was either absent or present at low cellular concentrations only. The low PUB/PEB ratio is consistent with the underwater light spectrum in the Baltic Sea. Near the Gulf of Finland, where the strains were isolated, the underwater light spectrum is dominated by green light (Stomp *et al.*, 2007a; Haverkamp *et al.*, 2008). Thus, our findings in the Baltic Sea add to earlier observations that picocyanobacteria of coastal seas typically have low PUB/PEB ratios (Olson *et al.*, 1990; Lantoine and Neveux, 1997; Wood *et al.*, 1998; Scanlan, 2003).

The use of 16S rRNA-ITS and *cpcBA* sequences as molecular markers for the phylogenetic relationships among *Synechococcus* strains allowed us to obtain further insight into the diversity of picocyanobacteria in the Baltic Sea. In a previous study, these loci were used to investigate the diversity within clone libraries constructed from environmental DNA (Haverkamp *et al.*, 2008). Several groups of picocyanobacteria that appeared in the clone libraries were not found among the cultured isolates, and vice versa (Table 2; Supplementary Figure SM5). This showed 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 observed picocyanobacterial diversity in the Baltic Sea considerably.

Analysis of the *cpcBA* operon largely separated the PE- and PC-rich strains in the BSea group. Hence, although PE- and PC-rich *Synechococcus* strains possess similar 16S rRNA-ITS sequences, they have quite different pigmentation (phycobilin) genotypes. Genetic differentiation between PE- and PC-rich strains by the *cpcBA* operon is ecologically highly relevant, because *Synechococcus* strains with different pigmentation are adapted to different underwater light climates (Pick, 1991; Vörös *et al.*, 1998; Wood *et al.*, 1998; Stomp *et al.*, 2004; Six *et al.*, 2007; Stomp *et al.*, 2007a; Haverkamp *et al.*, 2008).

Microdiversity, recombination and endemism

The results suggest that the diversity of picocyanobacteria in the Baltic Sea is substantially higher than that whichwe detected from our clone libraries and isolated strains. Rarefaction analysis using all known picocyanobacterial *cpcBA* sequences shows that the diversity within the picocyanobacterial *cpcBA* operon is not levelling off, even not at the lower similarity levels (Table 2). One possible explanation for the high microdiversity of *Synecho-*

coccus strains in the Baltic Sea is that the *cpcBA* operon evolves rapidly, with the third codon position showing many synonymous mutations (Haverkamp *et al*., 2008). The high picocyanobacterial diversity could also be caused by high recombination rates 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. ML 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 neighbour-net analysis, which showed high ambiguity within the phylogenetic signal. This indicates that evolutionary processes within the *cpcBA* operon cannot be easily resolved by phylogenetic analysis. This 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 recent genome analysis of several marine *Synechococcus* strains, which indicated that genes encoding PC and PE I and II show different evolutionary relationships in comparison with genes of the core genome such as the allo-PC gene or the ribosomal regions (Six *et al*., 2007).

Finally, neighbour-net analysis of the *cpcBA* phylogeny indicated that the PE-rich isolates of the BSea group are only distantly related to other known *Synechococcus* isolates (Figure 5). Hence, these strains may be endemic for the Baltic Sea with its unique environmental conditions. Studies on large ecosystems, such as Lake Superior and the Baltic Sea, indicate that locally adapted *Synechococcus* spp. inhabit these environments (Ivanikova *et al*., 2007; Haverkamp *et al*., 2008). This is consistent with the conclusions of Pommier *et al*. (2007), who found signals of endemism within the global bacterioplankton community. Therefore, we conclude that the high phylogenetic resolution provided by the *cpcBA* operon is useful to assess the microdiversity of *Synechococcus* strains in the Baltic Sea and other semienclosed ecosystems. Summarizing, our findings lend further support to the hypothesis (Stomp *et al*., 2007a; Haverkamp *et al*., 2008) that the underwater light spectra in waters of intermediate turbidity, which can be found in many mesotrophic lakes and coastal seas, such as the Baltic Sea, offer suitable niches for the widespread coexistence and extensive diversification of a colorful mixture of PE-rich and PC-rich picocyanobacteria.

Acknowledgements

We thank three anonymous referees for their very helpful and extensive comments and suggestions that improved an earlier version of this manuscript considerably. 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 ML comparisons. TH and LJS acknowledge support from the European Commission through the project MIRACLE (EVK3-CT-2002–00087). The research of JH was supported by the Earth and Life Sciences Foundation (ALW), which is subsidized by the Netherlands Organization for Scientific Research (NWO). This is publication 4444 of NIOO-KNAW.

References

Alonso C, Warnecke F, Amann R, Pernthaler J. (2007). High local and global diversity of *Flavobacteria* in marine plankton. *Environ Microbiol* **9**: 1253–1266.

Altschul SF, Madden TL, Zhang J, Lipman DJ, Schäffer AA, Zhang Z *et al*. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Bryant D, Moulton V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**: 255–265.

Chen F, Wang K, Kan JJ, Suzuki MT, Wommack KE. (2006). Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. *Appl Environ Microbiol* **72**: 2239–2243.

Crosbie ND, Pockl M, 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.

Ernst A, Becker S, Wollenzien UIA, Postius C. (2003). Ecosystem-dependent adaptive radiations of picocyanobacteria inferred from 16S rRNA and ITS-1 sequence analysis. *Microbiol SGM* **149**: 217–228.

Everroad RC, Wood AM. (2006). Comparative molecular evolution of newly discovered picocyanobacterial strains reveals a phylogenetically informative variable region of beta-phycoerythrin. *J Phycol* **42**: 1300–1311.

Felsenstein J. (1989). PHYLIP—phylogeny inference package (Version 3.2). *Cladistics* **5**: 164–166.

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. (2003). Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* **69**: 2430–2443.

Hall TA. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.

Haverkamp T, Acinas SG, Doeleman M, Stomp M, Huisman J, Stal LJ. (2008). Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ Microbiol* **10**: 174–188.

Herdman M, Castenholz RW, Iteman I, Waterbury JB, Rippka R. (2001). The Archaea and the deeply branching and phototrophic bacteria. In: Boone DR, Castenholz RW (eds). *Bergey's Manual of Systematic Bacteriology*, 2nd edn. Springer Verlag: Heidelberg. pp 493–514.

Huelsenbeck JP. (1995). Performance of phylogenetic methods in simulation. *Syst Biol* **44**: 17–48.

Huson DH, Bryant D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**: 254–267.

Ivanikova NV, Popels LC, McKay RM, Bullerjahn GS. (2007). Lake Superior supports novel clusters of cyanobacterial picoplankton. *Appl Environ Microbiol* **73**: 4055–4065.

Kisand V, Wikner J. (2003). Combining culture-dependent and -independent methodologies for estimation of richness of estuarine bacterioplankton consuming riverine dissolved organic matter. *Appl Environ Microbiol* **69**: 3607–3616.

Lantoine F, Neveux J. (1997). Spatial and seasonal variations in abundance and spectral characteristics of phycoerythrins in the tropical northeastern Atlantic Ocean. *Deep-Sea Res I* **44**: 223–246.

McGinnis S, Madden TL. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**: W20–W25.

Olson RJ, Chisholm SW, Zettler ER, Armbrust EV. (1990). Pigments, size, and distribution of *Synechococcus* in the North Atlantic and Pacific Oceans. *Limnol Oceanogr* **35**: 45–58.

Pick FR. (1991). The abundance and composition of freshwater picocyanobacteria in relation to light penetration. *Limnol Oceanogr* **36**: 1457–1462.

Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P *et al.* (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16**: 867–880.

Rippka R, Deruelles J, Waterbury JB. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111**: 1–61.

Robertson BR, Tezuka N, Watanabe MM. (2001). Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content. *Int J Syst Evol Microbiol* **51**: 861–871.

Rocap G, Distel DL, Waterbury JB, Chisholm SW. (2002). Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68**: 1180–1191.

Scanlan DJ. (2003). Physiological diversity and niche adaptation in marine *Synechococcus*. *Adv Microb Physiol* **47**: 1–64.

Schloss PD, Handelsman J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501–1506.

Schmidt TM, Delong EF, Pace NR. (1991). Analysis of a marine picoplankton community by 16S ribosomal-RNA gene cloning and sequencing. *J Bacteriol* **173**: 4371–4378.

Sieracki ME, Viles CL, Webb KL. (1989). Algorithm to estimate cell biovolume using image analyzed microscopy. *Cytometry* **10**: 551–557.

Six C, Thomas JC, Garczarek L, Ostrowski M, Dufresne A, Blot N *et al.* (2007). Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative genomics study. *Genome Biol* **8**: R259.

Stomp M, Huisman J, de Jongh F, Veraart AJ, Gerla D, Rijkeboer M *et al.* (2004). Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature* **432**: 104–107.

Stomp M, Huisman J, Stal LJ, Matthijs HCP. (2007b). Colorful niches of phototrophic microorganisms shaped by vibrations of the water molecule. *ISME J* **1**: 271–282.

Stomp M, Huisman J, Vörös L, Pick FR, Laamanen M, Haverkamp T *et al.* (2007a). Colourful coexistence of red and green picocyanobacteria in lakes and seas. *Ecol Lett* **10**: 290–298.

Swofford DL. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods) Version 4*. Sinauer Associates: Sunderland, Massachusetts.

Tamura K, Dudley J, Nei M, Kumar S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Vörös L, Callieri C, Balogh KV, Bertoni R. (1998). Freshwater picocyanobacteria along a trophic gradient and light quality range. *Hydrobiologia* **370**: 117–125.

Wood AM, Phinney DA, Yentsch CS. (1998). Water column transparency and the distribution of spectrally distinct forms of phycoerythrin-containing organisms. *Mar Ecol Prog Ser* **162**: 25–31.

Xia X, Xie Z. (2001). DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* **92**: 371–373.

Zaballos M, Lopez-Lopez A, Ovreas L, Bartual SG, D'Auria G, Alba JC *et al.* (2006). Comparison of prokaryotic diversity at offshore oceanic locations reveals a different microbiota in the Mediterranean Sea. *FEMS Microbiol Ecol* **56**: 389–405.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)