Phylogeography and pigment type diversity of Synechococcus cyanobacteria in surface waters of the northwestern Pacific Ocean

Xiaomin Xia, Frédéric Partensky, Laurence Garczarek, Koji Suzuki, Cui Guo, Shun Yan Cheung, Hongbin Liu

Phylogeography and pigment type diversity of *Synechococcus* cyanobacteria in surface waters of the northwestern Pacific Ocean

Xiaomin Xia[1], Frédéric Partensky[2], Laurence Garczarek[2], Koji Suzuki[3], Cui Guo[1], Shun Yan Cheung[1], Hongbin Liu[1]*

1. **Division of Life Science, The Hong Kong University of Science and Technology, Hong Kong**
2. **Sorbonne Universités, Université Paris 6, CNRS UMR 7144, Marine Plankton Group, MaPP team, CS 90074, Station Biologique, 29688 Roscoff Cedex, France**
3. **Faculty of Environmental Earth Science, Hokkaido University/JST- CREST, Japan**

Running Title: *Synechococcus* community diversity in the northwestern Pacific Ocean

*Corresponding author: E-mail liuhb@ust.hk; Tel. (+852) 23587341; Fax (+852) 23581559.

Accepted in Environmental Microbiology on Sept. 20, 2016

Summary

The widespread unicellular cyanobacteria *Synechococcus* are major contributors to global marine primary production. Here we report their abundance, phylogenetic diversity (as assessed using the RNA polymerase gamma subunit gene *rpoC1*) and pigment diversity (as indirectly assessed using the laterally transferred *cpeBA* genes, encoding phycoerythrin-I) in surface waters of the northwestern Pacific Ocean, sampled over nine distinct cruises (2008-2015). Abundance of *Synechococcus* was low in the subarctic ocean and South China Sea, intermediate in the western subtropical Pacific Ocean, and the highest in the Japan and East China seas. Clades I and II were by far the most abundant *Synechococcus* lineages, the former dominating in temperate cold waters and the latter in (sub)tropical waters. Clades III and VI were also fairly abundant in warm waters, but with a narrower distribution than clade II. One type of chromatic acclimater (3dA) largely dominated the *Synechococcus* communities in the subarctic ocean, while another (3dB) and/or cells with a fixed high phycourobilin to phycoerythrobilin ratio (pigment type 3c) predominated at mid and low latitudes. Altogether, our results suggest that the variety of pigment content found in most *Synechococcus* clades considerably extends the niches that they can colonize and therefore the whole genus habitat.

Key words: cyanobacteria, *Synechococcus*, *rpoC1*, western Pacific Ocean, marginal sea, pigment type, genetic diversity**.**

Abbreviations: phycocyanobilin (**PCB**); phycoerythrobilin (**PEB**); phycourobilin (**PUB**); type IV chromatic acclimaters (**CA4**); photosynthetically available radiation (**PAR**); the limit of detection (**LOD**); 16S–23S internal transcribed spacer region (**ITS**); the operational taxonomic unit (**OTU**); Non-metric multidimensional scaling analysis (**NMDS**); redundancy analysis (**RDA**).

Introduction

Marine unicellular cyanobacteria belonging to the *Synechococcus* genus are ubiquitously distributed in the euphotic layer of the ocean and play a key role in the marine ecosystems in terms of global carbon biomass (Garcia-Pichel *et al.*, 2003; Buitenhuis *et al.*, 2012) and net primary production (Liu *et al.*, 1998; Flombaum *et al.*, 2013). *Synechococcus* communities are both phenotypically and phylogenetically diverse. *Synechococcus* cells harvest light using large complexes called phycobilisomes, constituted by a core of allophycocyanin surrounded by six radiating rods with highly variable phycobiliprotein and chromophore composition (Six *et al.*, 2007). Based on the different phycobiliprotein composition of phycobilisome rods, *Synechococcus* strains can be divided into three main pigment types: type 1 containing only phycocyanin (PC), type 2 containing both PC and phycoerythrin-I (PEI) and type 3 possessing PC, PEI and phycoerythrin-II (PEII). PEI and PEII have different amino acid sequence and bilin composition and content, with the latter always binding both phycoerythrobilin (PEB) and phycourobilin (PUB), whereas the former either bind only PEB or both PEB and PUB (Ong and Glazer, 1991; Six *et al.*, 2007). In terms of chromophorylation, phycobilisomes of pigment type 1 bind only phycocyanobilin (PCB), type 2 bind both PCB and PEB, while type 3 bind PCB, PEB and PUB. Pigment type 3 *Synechococcus* can be further divided into type 3a (PUB/PEB<0.6), 3b ($0.6 \leq$ PUB/PEB<1.6), 3c (PUB/PEB$\geq$1.6) and type 3d/e (variable PUB/PEB), according to the *in vivo* PUB/PEB ratio of whole cells, as assessed using the ratio of the fluorescence excitation peaks of these two chromophores, i.e. usually $Ex_{495\ nm}$:$Ex_{550\ nm}$ (Everroad *et al.*, 2006; Six *et al.*, 2007, Everroad and Wood, 2012; Humily *et al.*, 2013). Strains displaying either high (type 3d) or low (type 3e) amplitude of variation of their PUB/PEB ratio

in response to change in light quality from blue to green light (and vice versa) are also called type IV chromatic acclimaters (CA4). Genes involved in the CA4 process are gathered into a small genomic island with two possible configurations, CA4-A and CA4-B (Humily *et al.*, 2013). The chromatically acclimating strains that possess a CA4-A or a CA4-B island type are called 3dA or 3dB, respectively, but it is noteworthy that a number of strains that possess such an island do not actually chromatically acclimate and therefore display a fixed PUB/PEB ratio, possibly due to a dysfunction of the CA4 regulatory machinery.

The distribution of pigment types has been studied in the marine environment using various approaches, including epifluorescence microscopy (Wang *et al.,* 2011), flow cytometry (Olson *et al.,* 1988), spectrofluorimetry (Lantoine and Neveux, 1997; Wood *et al.,* 1998) and sequencing of clone libraries of *cpcBA* and *cpeBA* operons, encoding alpha and beta subunits of PC and PEI, respectively (Liu *et al.,* 2014; Chung *et al.,* 2015). Indeed, in sharp contrast with phylogenies obtained from housekeeping genes such as 16S rRNA or *rpoC1* genes, phylogenies made using such phycobiliprotein-encoding genes reflect better pigment content than vertical phylogeny (Six *et al*., 2007; Everroad *et al*., 2012; Humily *et al*., 2014). While the characterization of cultured *Synechococcus* isolates showed that different pigment types can collect different wavelengths of light (Olson *et al.,* 1988; Olson *et al.,* 1990; Stomp et al., 2004; Six *et al*.,2007), field studies suggest that these differences can allow them to colonize distinct light quality niches of the marine environment (Wood, 1985; Wood *et al*., 1999; Stomp *et al*., 2007; Liu *et al*., 2014). Pigment type 1 *Synechococcus* harvest optimally red-orange light and dominate in turbid estuarine waters (Wang *et al*., 2011; Liu *et al*., 2014), whereas type 2 mainly absorb yellow-green light and preferentially thrive in turbid coastal or continental shelf waters

(Wood *et al.*, 1998). The different pigment subtypes within type 3 would mainly occur in mesotrophic or oligotrophic oceanic waters where green or blue light predominate, respectively (Olson *et al.*, 1988; Wood *et al.*, 1998). Both cultivation-dependent and independent methods used to study the pigment diversity of *Synechococcus* communities have shown that multiple pigment types co-occur in marine surface waters (Choi and Noh, 2009; Haverkamp *et al.*, 2009; Humily *et al.*, 2014; Larsson *et al.*, 2014).

The vertical phylogeny of the marine *Synechococcus* group has also been extensively studied. Based on the 16S rRNA gene marker, marine *Synechococcus* have been classified into three phylogenetic subclusters, 5.1, 5.2 and 5.3 (hereafter S5.1, S5.2 and S5.3; Dufresne *et al.*, 2008). S5.1, which is more diverse than the other two subclusters, contains at least sixteen clades (Ahlgren and Rocap, 2012). S5.1 and S5.3 are mainly composed of pigment types 2 and 3, except the euryhaline clade VIII (within S5.1), which so far only comprises strains belonging to pigment type 1. In contrast, S5.2 contains pigment types 1 and 2 (e.g., strains CB0101 and CB0205; Chen *et al.*, 2006). Studies of the global distribution of *Synechococcus* community composition have revealed a clear spatial partitioning for the main *Synechococcus* lineages (Zwirglmaier *et al.*, 2008; Huang *et al.*, 2012; Mazard *et al.*, 2012; Sohm *et al.*, 2015; Farrant *et al.*, 2016). Indeed, in agreement with the different thermal growth range and preference of representative strains (Pittera *et al.*, 2014), clade II was found to be prevalent in warm subtropical and tropical waters, while clades I and IV dominate in cold waters. The relative abundance of *Synechococcus* lineages also undergo seasonal variations (Fuller *et al.*, 2003; Post *et al.*, 2011; Chung *et al.*, 2015). For instance, clade III often occurs in summer in warm, stratified surface waters, whereas clade XII are generally prevalent in fall and winter (Post *et*

121 *al.*, 2011). Although 16S rRNA gene was the first phylogenetic marker used to define
122 *Synechococcus* lineages (Urbach *et al.*, 1998; Fuller *et al.*, 2003), some of these lineages cannot
123 be discriminated by this marker, due to its too low taxonomic resolution. Therefore, several
124 other markers such as the 16S–23S internal transcribed spacer region (ITS; Ahlgren and Rocap,
125 2012) or the single copy genes *rpoC1* (Toledo and Palenik, 1997), *ntcA* (Penno *et al.*, 2006)
126 and *petB* (Mazard *et al.*, 2012) have been applied to study the genetic diversity of
127 *Synechococcus* at higher resolution.

128 The western Pacific Ocean, an area extending from 15 to 55 °N, exhibits a wide variety
129 of hydrographic conditions with surface water temperature ranging from > 29 °C throughout
130 the year in the tropical ocean to < 10 °C in the subarctic ocean, providing a comprehensive
131 framework to study the effects of physico-chemical parameters on *Synechococcus* community
132 structure. This area encompasses five marginal seas, including the South China Sea, the East
133 China Sea, the Japan Sea, the Sea of Okhotsk and the Bering Sea, located along the
134 northwestern boundary of the Pacific Ocean, and altogether covering the tropical, temperate
135 and frigid zones (Fig. 1A). The warm current Kuroshio (Sawada and Handa, 1998) and the cold
136 current Oyashio (Qiu, 2001) as well as their extension (the Northern Pacific current; Cummins
137 and Freeland, 2007) also create some unique environments in the western Pacific Ocean. For
138 example, the Tokara Strait where the Kuroshio Current flows out of the East China Sea to
139 Pacific Ocean, has water temperature higher than the adjacent regions and does not exhibit
140 strong seasonal variations (Nagata and Takeshita, 1985), whereas the western subtropical
141 Pacific (East of Taiwan) is part of the western Pacific warm pool, a typical oligotrophic
142 environment with low nutrient concentration. As some seasonal variations in the abundance of

Synechococcus cells have been reported in the South China Sea (Xia *et al.*, 2015a) and East China Sea (Guo *et al.*, 2014), the present study mainly focused on late spring and summer (with the exception of Tokara Strait samples that were sampled in fall), in order to lower potential effects of seasonal variations on data interpretation. Besides covering a wide temperature and nutrient gradients, the study region also encompasses a large variety of optically distinct water bodies, from turbid coast to transparent open ocean waters. Therefore, this region likely comprises most of the niche types of *Synechococcus* that are reported so far in global ocean. Studying concomitantly the phylogenetic and pigment type composition of *Synechococcus* communities across such a wide range of environments allowed us to reveal the latitudinal distribution patterns of different clades and pigment types, to assess the relative importance of temperature and other environmental factors (in particular light quality) in determining niche partitioning of *Synechococcus* populations, and to unveil some of the strategies developed by *Synechococcus* to adapt to these different niches.

Results

Synechococcus is widespread in surface waters of the northwestern Pacific Ocean

Flow cytometry analyses showed that *Synechococcus* ubiquitously occurred in the surface waters (5-10 m depth) of northwestern Pacific Ocean (Fig. 1B). During the sampling periods, which are mainly late spring to summer, *Synechococcus* abundance was low in the subarctic waters (Sea of Okhotsk, Bering Sea and western subarctic Pacific Ocean; 0.7-6.7×10^3 cells mL^{-1}, average 1.7×10^3 cells mL^{-1}) and South China Sea (2.5×10^3-3.9×10^4 cells mL^{-1}, average 7.1×10^3 cells mL^{-1}), slightly higher in western subtropical Pacific Ocean (2.5×10^3-

165 3.7×10^4 cells mL^{-1}, average 1.2×10^4 cells mL^{-1}), and the highest in the East China Sea (8.0×10^3-
166 2.8×10^5 cells mL^{-1}, average 7.1×10^4 cells mL^{-1}) and Japan Sea (1.9×10^3-1.1×10^5 cells mL^{-1},
167 average 6.1×10^4 cells mL^{-1}). In the Tokara Strait, a region located South of Japan that was
168 sampled in November 2012, the abundance of *Synechococcus* was fairly low (1.3-2.2×10^4 cells
169 mL^{-1}; average 1.7×10^4 cells mL^{-1}) but comparable with that in open ocean waters of the East
170 China Sea collected in August 2009, suggesting that this was not due to seasonal variations
171 (Fig. 1B). The abundances of *Synechococcus* were the most variable in the East China Sea,
172 with the highest ones occurring in the northern part and in the mid-shelf region and the lowest
173 ones in the estuary of Changjiang River. In the South China Sea, the abundance of
174 *Synechococcus* was slightly higher in coastal than in open ocean waters. In the subarctic ocean,
175 *Synechococcus* abundance was only around 5.0×10^2 cells mL^{-1} in the Bussol Strait connecting
176 the Sea of Okhotsk and the Pacific Ocean, where a strong vertical mixing occurred. This was
177 the lowest *Synechococcus* abundance recorded in all regions investigated in this study (Fig.
178 1B).

179 Altogether, the abundance of *Synechococcus* in surface water showed a weak positive
180 correlation with temperature and a weak negative correlation with PO_4^{3-} (Table S1). However,
181 when looking at individual geographic regions, nutrients as well as temperature were often
182 found to play a strong role in influencing the local abundance of *Synechococcus*. For example,
183 cell density in the western subtropical Pacific Ocean was found strongly affected by PO_4^{3-},
184 whereas TIN, PO_4^{3-}, as well as temperature had a positive influence in the subarctic ocean. In
185 the South China Sea, which displays a fairly low temperature with regard to the western
186 subtropical Pacific Ocean, potentially due to the occurrence of a local coastal upwelling, we

observed a weak negative relationship between temperature and *Synechococcus* abundance. In contrast, no significant correlation was observed between *Synechococcus* abundance and salinity, even in the East China Sea region that was influenced by local river input (Table S1).

The different pigment types of *Synechococcus* colonize different niches in the environment

The diversity of *Synechococcus cpeBA* operon in each community was estimated using the Shannon diversity index (Fig. S1). The Shannon diversity index of the 12 *Synechococcus* communities ranged from 0 to 2.18. WSAP-C5, which was collected from the western subarctic Pacific Ocean, had the lowest diversity, while TS-ST4, collected from the Tokara Strait, had the highest diversity. Altogether, the diversity of *Synechococcus cpeBA* operon was higher in tropical and subtropical waters than in temperate waters (Fig. S1).

Phylogenetic analysis of the *cpeBA* operon sequences allowed us to separate four groups of pigment types (Fig. 2): 2, 3a, 3dA and the combination of 3c and 3dB, since the latter two types cannot be discriminated using this marker gene (Humily *et al*., 2014). It is worth noting that the only two strains known so far to exhibit a pigment type 3b in culture (WH8103 and WH8109) are both genetically similar to 3dB strains and possess a complete CA4-B island, but they have lost their ability for chromatic acclimation (Humily *et al.*, 2013). Thus, pigment types 3b and 3dB also cannot be differentiated based on phylogenies using *cpeBA* (Humily *et al.*, 2014). With these caveats in mind, our phylogenetic analyses showed that all pigment types detectable with this marker were present in the northwestern Pacific Ocean, but their relative abundance exhibited large geographical variations (Fig. 3). *Synechococcus* pigment type 2 was

by far the least abundant, mainly occurring at ECS-KP01, a station influenced by a river plume. Pigment type 3a was abundant in the Tokara strait and dominant in the whole East China Sea region (Fig. 3). The group composed by type 3c and/or type 3dB was largely dominant at both stations of the South China Sea as well as at the southernmost station of the Northern Pacific Current. Most strikingly, chromatic acclimaters of the 3dA type constituted almost the sole pigment type detected in the western subarctic Pacific Ocean and Bering Sea, but occurred in mixture with 3dB/3c and 3a types at the northernmost station of Northern Pacific Current.

Synechococcus genetic diversity is much higher in the subtropics than in the subarctic ocean

In this study, the *rpoC1* gene of 33 samples from the northwestern Pacific Ocean (Fig. 1A) were sequenced. After cleaning, 1,739 high-quality sequences of each sample remained for further analysis. The Shannon diversity index of *Synechococcus* communities ranged from 0.99 to 4.34 (Fig. S2). Generally, the diversity of *Synechococcus* in summer was low in the subarctic ocean (Bering Sea, Sea of Okhotsk and western subarctic Pacific Ocean), slightly higher in the South China Sea, and the highest in the East China Sea (Fig. S2). In particular, ECS-DH13 and ECS-PN05 showed very high diversity index values. The Shannon diversity based on the *rpoC1* sequences was systematically higher than that based on the *cpeBA* sequences at all studied stations.

In the western Pacific Ocean, during summer, the *Synechococcus* community composition at sites located above and below 40ºN (Groups A and B on Fig. 4A, respectively) were significantly different (P=0.001, One-way Analysis of Variance). Sites NPC-J6, JS-TE3,

JS-JY1, and JS-SH2, which lie between latitudes 34ºN and 40ºN did not fall in any of these groups, indicating that the *Synechococcus* community compositions at those sites were unique and distinct from all other stations. Since all Japan Sea samples were collected in spring, we cannot exclude that their specificity results in part from seasonal variations. In the case of the Northern Pacific Current cruise, the two southernmost samples from the (NPC-K1 and A1), a region also sampled in spring, fell close to Tokara strait samples collected at the same latitude but in winter. Yet the latter set of samples fell within group B, which mainly gathers low latitude samples collected during summer, suggesting that latitude has a stronger effect on genetic diversity than season. To better split *Synechococcus* communities falling within group B (average similarity: 29.31%), we did a further NMDS analysis specifically on this group (Fig. 4B). In the East China Sea, *Synechococcus* communities formed two major sub-groups composed of samples collected below and above 32ºN, respectively (Fig. 4B). Similar result was observed in the North Pacific Current (Fig. 4A). The coastal station ECS-KP01, which is strongly influenced by the river discharge, did not group together with other East China Sea samples.

More than 80% of reads in each sample were taxonomically assigned to known lineages using the reference database, except stations SCS-SEATS, SCS-LE09 and WSTP-ST1, where a higher number of unknown sequences were detected, potentially corresponding to novel clades (Fig. S3). All three *Synechococcus* subclusters were detected but S5.1 was by far the most abundant and widespread and constituted the only subcluster detected in the subarctic ocean (Fig. S3). S5.2 was only found in 13 samples, and its relative abundance was lower than 1% in 8 of them. The highest abundance of this subcluster (15.5%) was detected at the coastal

253 ECS-KP01 station. S5.3 was widespread but generally occurred at low abundance in the
254 tropical and subtropical ocean and was virtually absent in cold waters of the Japan Sea, Sea of
255 Okhotsk, western subarctic Pacific Ocean and Bering Sea. Yet, its abundance was fairly high
256 in the Tokara Strait, especially at TS-ST5 (Figs. 5 and S3).

257 Altogether, 17 clades within S5.1 were detected in the western Pacific Ocean with clades
258 I, II, III and VI being the most abundant (Fig. 5). Clade I mainly occurred in the subarctic ocean,
259 while clade II widely dominated in tropical and subtropical areas. Compared to clade II, clade
260 III also occurred in warm waters, but with a lower abundance and a narrower distribution (Figs.
261 5 and S3). Clade VI was also widely distributed in warm East China Sea (DH04 and HH12)
262 and TS waters that are influenced by the Kuroshio Current. Clades VII and WPC2 exhibited a
263 similar niche as S5.3, all displaying a wider distribution than clade VI. Clades Miyav and IX
264 co-occurred with S5.2, mainly in waters influenced by a nutrient-rich and turbid river plume.
265 WPC1 was mainly found in the East China Sea, with a similar distribution to clade III. Clade
266 XV co-occurred with clade XVI and CRD1, especially in the Northern Pacific Current.

267 RDA analysis showed that temperature, which explained 56.0 % of the variance of
268 *Synechococcus* community, was by far the most important factor determining the distribution
269 of *Synechococcus* lineages in the northwestern Pacific Ocean (Fig. 6A). Clade I and IV relative
270 abundances were associated both with low temperature and high nutrients, while all other
271 clades showed reverse correlations with these parameters (Fig. 6A). This is consistent with the
272 occurrence of these two clades being almost restricted to high latitudes (> 35°N), where the
273 surface water temperature was lower than 12 ℃, while clade II often accounted for more than
274 60 % of the sequences from 18° N to 34° N (Fig. 7A). Other important lineages, i.e. clades III,

VI and S5.3, were only abundant between 25ºN and 35ºN (Fig. 7B).

Focusing on stations with temperature higher than 25 ℃ then allowed us to better discriminate the other important parameters driving the distribution of all lineages except clades I and IV (Fig. 6B). Clade IX, Env-Miyav and S5.2 were positively correlated to TIN and phosphate concentrations, consistent with their preference for river plume influenced waters (Figs. 6 and S3). Clades II and III also appeared to be well separated on correlation biplots (Fig. 6B). However, the parameters explaining their respective distribution remain unclear even though salinity could partially explain the observed differences, clade III being most abundant in the East China Sea (Fig. S4), which displayed the lowest salinity of the studied area (Table S2).

Unveiling *Synechococcus* diversity within clade I in the western Pacific Ocean

Although *Synechococcus* clade I is generally considered as typical of cold, mesotrophic and eutrophic waters (Zwirglmaier *et al.*, 2008), members of this clade were also detected at low abundance in warm waters of East China Sea, TS and Northern Pacific Current (Fig. 8). Phylogenetic analysis of clade I *rpoC1* sequences showed that OTUs, which dominated in the subarctic ocean (Bering Sea, Sea of Okhotsk and western subarctic Pacific Ocean) clustered together. In contrast, OTUs occurring at lower latitude (East China Sea, Tokara Strait, Japan Sea and Northern Pacific Current) could be split into five additional subclades, displaying slightly different distribution patterns. Thus, one subclade is truly psychrophilic as it mainly dominated in subarctic oceanic waters, where the surface water temperature was lower than 12 ℃, while other subclades occur in warm waters with temperature ranging from 15 to 29 ℃.

297 Fig. 8 also shows that the diversity of *Synechococcus* clade I was the highest between 34ºN
298 and 40ºN.
299
300 **Linking *Synechococcus* genetic diversity and pigment content.**
301 In order to examine the links between *Synechococcus* phylogeny, as assessed using the
302 *rpoC1* marker, and pigment content, as indirectly assessed using the *cpeBA* operon, a Spearman
303 rank correlation analysis was performed over the whole Northwestern Pacific Ocean dataset
304 (Fig. S5). It showed that pigment type 2 (no PUB) was strongly correlated with clades IX and
305 Env-Miyav, while type 3a (low PUB:PEB) was associated with clades III, V and WPC1.
306 Pigment type 3dA (one of the two types of chromatic acclimaters) was strongly positively
307 correlated with clade I but anti-correlated with clade II. At last, the two indistinguishable
308 pigment types 3c (high PUB:PEB) and 3dB (the second chromatic acclimater type) were
309 associated with clades II and UC-A. Thus, even if pigment types are not restricted to single
310 clades, they seem to be preferentially found in certain clades in the field.
311
312
313 **Discussion**
314 **Factors controlling the abundance of *Synechococcus* in the northwestern Pacific Ocean**
315 This study, which analyzed surface samples collected from nine distinct cruises in the
316 northwestern Pacific and its marginal seas, provides a comprehensive overview of the
317 variability of *Synechococcus* abundance in this vast oceanic region and novel insights on how
318 it relates with environmental factors. While the *Synechococcus* abundance was high in the

subtropical and temperate areas (25-40°N), especially near the coasts (East China Sea and Japan Sea), it was fairly low in the tropical (South China Sea) and subarctic regions (Figs.1B and 7A). Using a parametric regression model, a recent study also predicted strong variations of *Synechococcus* abundance with latitude, but when averaged over a year, the abundance peak was located around 45°N (Flombaum *et al.*, 2013), i.e. at a significantly higher latitude than observed here (around 33.5°N; Fig. 1B). This difference may arise in part from seasonal variations in *Synechococcus* abundance (see e.g. Fig. S5 in Flombaum *et al.*, 2013) but also from the fact that latitudinal variations of abundance could significantly differ between the marginal seas studied here and the central part of the northwestern Pacific Ocean, and this local distribution pattern may have been missed by the global model developed by Flombaum and coworkers. Indeed, the highest cell densities reported here were all recorded in coastal seas (Fig. 1B), with a maximal abundance of 2.8×10^5 *Synechococcus* cells mL^{-1} in the East China Sea in summer. This value is one order of magnitude higher than those reported during summer 1998 in the same area (Jiao *et al.*, 2005), but is comparable to those reported in other nutrient-rich areas around the world, such as the Bedford Basin (Li, 1998) and the Martha Vineyard's coastal observatory in summertime (Hunter-Cevera *et al.*, 2015), as well as the Arabian Sea (Liu *et al.*, 1998) or areas influenced by local upwelling (Partensky *et al.*, 1996). It is however much lower than those observed in the Costa Rica dome, where record abundances between 1.2×10^6 and 3.7×10^6 cells mL^{-1} have been reported (Saito *et al.*, 2005).

A number of previous studies have suggested that the net abundance of *Synechococcus* is mainly controlled by temperature and nutrients (Agawin *et al.*, 2000; Zwirglmaier *et al.*, 2008; Tai and Palenik, 2009). Here we found that when considering our whole dataset on

surface waters of the western North Pacific Ocean, *Synechococcus* abundance was only weakly correlated with temperature and phosphate availability and not correlated with TIN availability (Table S1). Yet, this overall trend clearly masks local disparities. As for the first parameter, temperature seems to be more influential to *Synechococcus* abundance in low temperature subarctic ocean than tropical/subtropical warm waters. This is in agreement with a previous study showing that *Synechococcus* abundance is positively correlated to temperature only below 14 ℃ (Li, 1998). Although Flombaum and coworkers predicted that the abundance of *Synechococcus* cells in the world ocean will increase by 14% at the end of this century, due to the global rise in sea surface temperature (Flombaum *et al.*, 2013), they also foresaw that this increase will not equally affect all latitudes. Our results indeed suggest that in the northwestern Pacific Ocean, effects of global change should be much more significant in the subarctic region, where both surface temperature and *Synechococcus* abundance are low, than in temperate, tropical or subtropical waters.

As concerns nutrient concentrations, our statistical analyses showed that *Synechococcus* abundance was positively influenced by nutrient availability in the subarctic ocean but not in the tropical/subtropical waters (Table S1). *Synechococcus* abundance in the latter regions was higher in coastal and river plumes influenced waters than in oligotrophic oceanic waters and low salinity estuarine waters (Fig. 1B). This is in agreement with the dome-shaped distribution proposed by Liu *et al.,* (1998), which suggested that *Synechococcus* grow best in waters with intermediate level of nitrate. Consistently, Chung and coworkers, who also studied the distribution of *Synechococcus* in the East China Sea observed during non-flooding summer the occurrence of a bloom of PE-rich *Synechococcus* on the outer boundaries of the Changjiang

River in water with a salinity comprised between 31 and 32 ppt (Chung *et al.*, 2014).

Niche partitioning of marine *Synechococcus* pigment types

The wide diversity of pigment types that we observed in the northwestern Pacific Ocean translates the large variety of light environments encountered in this vast oceanic region, from turbid estuaries to transparent open ocean waters (Fig. 3). Interestingly, while different *Synechococcus* pigment types co-occurred within most samples, one phenotype generally predominated. This observation supports the hypothesis that the spectral properties of seawater exert a strong selective pressure on *Synechococcus* populations, favoring the pigment type possessing phycobilisomes with light absorption characteristics matching at best the local spectrum of photosynthetically available radiation (PAR). While low PUB:PEB cells (type 3a) dominated the population all over the East China Sea, *Synechococcus* cells lacking PUB (type 2) were only abundant in a low transparency station (St. ECS-KP01, see Fig. 2 of Chung *et al.*, 2014). This indicates that the latter pigment type is well adapted to harvest light in these fairly turbid waters (Olson *et al.*, 1988; Wood *et al.*, 1998), where the PAR spectrum is likely shifted towards yellow/yellow-green light due to organic matter in suspension (Kirk, 1994), while pigment type 3a can stand wider PAR spectra, extending from blue-green to yellow-green (see e.g., Six *et al.*, 2007 for representative absorption spectra of pigment types 2 and 3). It is noteworthy that although analysis of the *cpeBA* operon diversity does not allow to detect *Synechococcus* type 1 cells since they lack phycoerythrin, this pigment type was probably also present at station ECS-KP01, as these phycocyanin-rich *Synechococcus* were recently shown to predominate in the highly turbid diluted waters of the Changjiang River (Chung *et al.*, 2014).

As mentioned above, type 3dB cells, (i.e. chromatic acclimaters possessing a CA4-B genomic island; Humily *et al.*, 2013), are not distinguishable from type 3c (i.e. cells with a fixed high PUB:PEB ratio) based on analyses of the *cpeBA* operon. It is therefore not possible from our data to precisely assess the respective habitat of these two pigment types, though it appears that both are absent from turbid, coastal as well as high latitude waters. Previous studies using flow cytometry or spectrofluorometry have concluded that blue, oligotrophic areas are populated with type 3c cells (Olson *et al.*, 1988; Campbell and Iturriaga, 1988; Lantoine and Neveux, 1997; Wood *et al.*, 1999; Haverkamp *et al.*, 2009; Everroad and Wood, 2012). Yet, it is worth noting that these methods also cannot distinguish between type 3c and 3dB cells, since both these pigment types exhibit high PUB:PEB ratios in blue light (Humily *et al.*, 2013). In contrast, our study clearly indicated, for the first time, that type 3dA, i.e. chromatic acclimaters possessing a CA4-A genomic island, were by far the predominant pigment type in the subarctic ocean area. Due to the low angle of the sun at these latitudes, light intensity decreases rapidly with depth, resulting in an unstable light environment (Nosaka *et al.*, 2014). Thus, the ability to perform chromatic acclimation seemingly constitutes an advantage in environments with variable light conditions, consistent with a previous study showing the abundance of chromatic acclimaters in permanently mixed waters of the English Channel (Humily *et al.*, 2014).

Niche partitioning of *Synechococcus* clades

Synechococcus clades are known to partition more strongly along the horizontal scale than with depth (Zwirglmaier *et al.*, 2008; Choi and Noh, 2009; Sohm *et al.*, 2015), indicating that even though our study dealt only with populations from surface waters, the relative

abundance of clades that we observed at any given station is truly representative of the diversity of local *Synechococcus* populations. Several studies have suggested that clades I through IV are the most abundant clades at the global scale, with clades I and IV predominantly found in cold and temperate waters, while clades II and III prefer warmer waters (Zwirglmaier *et al.*, 2008; Mella-Flores *et al.*, 2011; Post *et al.*, 2011; Huang *et al.*, 2012). Here we show that clade I was by far the dominant clade in the subarctic ocean and temperate waters of Japan Sea and Northern Pacific Current, whereas clade II appears to be its counterpart in subtropical and tropical areas (Figs. 5 and S4), suggesting that these two dominant clades are mutually exclusive. In contrast, clade IV was scarce over all the studied area, including cold waters, in agreement with a previous report of low abundances of the latter clade in the Bering Sea (Huang *et al.*, 2012).

Interestingly, a number of clade I sequences was also retrieved from East China Sea, Tokara Strait and Northern Pacific Current (Fig. 8), consistent with previous studies reporting the occurrence of this clade as a minor component of *Synechococcus* communities in tropical and subtropical, warm waters (Fuller *et al.*, 2006; Ahlgren *et al.*, 2014). This wide distribution of clade I might be due in part to the large microdiversity existing within this clade. A single subclade encompassing eight OTUs (subclade I-F) dominated in the subarctic ocean cluster, while OTUs present in warm and temperate waters (East China Sea, Tokara Strait and Northern Pacific Current) appeared more diverse, clustering into 5 separate subclades (I-A through I-E; Fig. 8). Interestingly, members of all six subclades co-occurred at intermediate latitudes (Japan Sea and/or Northern Pacific Current), suggesting that the realized niches of warm- and cold-adapted clade I populations overlap in this transition area in which genetic exchange could

occur, e.g., by lateral gene transfer. The wider abundance and geographic distribution of clade I compared to clade IV could also possibly be explained by the fact that members of the former clade are genetically more versatile than members of the latter clade, as suggested by comparative genome analysis of strains representative of these two clades (Dufresne *et al.*, 2008; Scanlan *et al.*, 2009; Tai and Palenik, 2009)

Although clade III was locally abundant in warm waters (>25 ℃), it had a more limited distribution pattern than clade II and was only abundant in the nutrient-rich and/or low salinity waters of East China Sea (Figs. 5 and S4). This observation is in sharp contrast with the realized niche of this clade in the eastern basin of the Mediterranean Sea, a very oligotrophic, strongly phosphate-depleted, high-salinity area, where clade III constitutes the locally dominant *Synechococcus* taxon (Mella-Flores *et al.*, 2011; Farrant *et al.*, 2016). This suggests that, like clade I, clade III encompasses several subpopulations with very distinct nutrient and/or salinity preferenda, i.e. distinct ecotypes.

The fourth most abundant taxon in the northwestern Pacific was clade VI but its geographical distribution was limited to the Tokara strait and the central part of East China Sea (Figs. 5 and S4), suggesting that it has a narrow niche. Yet, correlation analyses did not allow us to clearly identify factors delineating its niche, and its overall distribution is also not well understood so far since counts of this clade have often been merged with those of the related clades V, VII and/or CRD1 (Zwirglmaier *et al.*, 2008; Huang *et al.*, 2012). Thus, more field and culture studies are clearly needed to characterize this group. A few other clades within S5.1 as well as members of S5.3 were also frequently encountered in northwestern Pacific Ocean warm waters but generally at lower relative abundances (Fig. 5). Clade WPC1, which was first

451 reported in the East China Sea and Japan Sea (Choi and Noh 2009), may co-occur with clade
452 III, whereas *Synechococcus* clades VII and S5.3 occur in waters with variable nutrients and
453 light supply (Fuller *et al.*, 2006; Post *et al.*, 2011). Although previous studies suggested that
454 clades XV and XVI have a global distribution (Ahlgren and Rocap, 2006; Sudek *et al.*, 2015),
455 we rather suggest that these two clades mainly occur between 30° and 35° N/S (Ahlgren and
456 Rocap, 2006; Huang *et al.*, 2012; Sudek *et al.*, 2015) and in upwelling regions (Sohm *et al.*,
457 2015), i.e. in areas characterized by nutrient-rich and intermediate environmental conditions.
458 At last, members of clades IX and Miyav as well as S5.2 showed similar niches, with an
459 abundance peak in the river plume influenced waters (Fig. 5). These taxa were previously found
460 to be abundant in the Hong Kong estuarine waters (Xia *et al.*, 2015b), suggesting that they have
461 a high nutrient requirement and are possibly halotolerant, as it is the case for S5.2 (Chen *et al.*,
462 2006; Dufresne *et al.*, 2008).
463
464 **The combination of genetic and pigment diversity contributes to *Synechococcus* ubiquity**
465 Our phylogenetic tree based on *cpeBA* operon sequences clearly grouped together
466 members of several distinct phylogenetic clades (Fig. 2). This is consistent with several
467 comparative phylogenetic analyses based on the one hand on phycobilisome rod genes and on
468 the other hand on housekeeping genes (ITS, ribosomal proteins, etc.) or allophycocyanin genes
469 that have suggested that the variety of pigment types that occurs among lineages of S5.1 results
470 from multiple lateral transfers of PE-encoding genes between *Synechococcus* lineages during
471 the evolution of this genus (Six *et al.*, 2007; Haverkamp *et al.*, 2008; Haverkamp *et al.*, 2009).
472 Another study has also evoked a specific loss of *mpeBA* genes in pigment type 2 strains as an

alternative hypothesis (Everroad and Wood, 2012). Whatever the origin of this discrepancy between *Synechococcus* clades and pigment types, it remains possible to assign with some confidence a pigment type to a specific clade in the field, whenever there is a concomitant dominance at one location of both one pigment type and one clade over all others. For instance, clade I populations from the subarctic North Pacific Ocean are clearly almost exclusively of type 3dA. This is consistent with the fact that all clade I strains able to chromatically acclimate sequenced so far possess a CA4-A island (Humily *et al.*, 2013). Similarly, the comparison of Figures 3 and 5 suggest that most cells at St. ECS-KP01 seemingly belong to clade Miyav and have phycobilisomes of type 3a, while those at St. SCS-LE04 predominantly belong to clade II and possess phycobilisomes of type 3c and/or 3dB (Figs. S5 and S6).

It also appears that members of *Synechococcus* communities belonging to a single clade can possess phycobilisomes of different types, consistent with observations on isolates (Fig. 2, but see also Six *et al.*, 2007; Haverkamp *et al.*, 2008; Haverkamp *et al.*, 2009; Everroad and Wood 2012). For instance, although the *Synechococcus* community at St. ECS-KP13 was largely dominated by clade II (Fig. 5), there were several co-occurring pigment types at this station: 3a, 3c and/or 3dB (Fig. 3). It is noteworthy though that the low proportion of pigment type 3dA sequences also found at this station is likely attributable to minor clades, possibly WPC2 or UC-A (Fig. 5), since all clade II and III strains able to chromatically acclimate sequenced thus far are of pigment type 3dB, that is possess a CA4-B island (Humily *et al.*, 2013).

In conclusion, although combining genetic and pigment diversity analyses has rarely been applied to field populations prior to this study, this approach clearly brings interesting new

perspectives on the extent of the flexibility of the *Synechococcus* genus as a whole with regard to environmental parameters. The highly variable phycobiliprotein composition and chromophore content of *Synechococcus* antenna complexes, and hence particularly wide PAR range in which cells of this genus can thrive, clearly confers this group a flexibility with regard to light quality that is unique among marine phytoplankters. Yet, the adaptive capacity of *Synechococcus* is not limited to the spectral properties of seawater, but also concerns other key parameters, including temperature, salinity and nutrient availability, for which adaptive specific traits have been selected by billions years of evolution (Dufresne *et al.*, 2008; Scanlan *et al.*, 2009; Pittera *et al.*, 2014). Altogether this adaptability explains the extraordinary ubiquity of *Synechococcus* not only in the northwestern Pacific Ocean but more generally in the marine environment. Our study provides some unprecedented general patterns of *Synechococcus* abundance, pigment diversity and clade composition in a key oceanic region, the northwestern Pacific Ocean. Despite the fact that most samples were collected during summer, we cannot rule out that some of the observed variability actually arises from temporal changes, since seasonal variations of *Synechococcus* abundance and community composition have been reported in some regions of the western Pacific Ocean (Xia *et al.* 2015a) and elsewhere (Tai and Palenik 2009). Thus, future studies are required to better decipher the effects of seasonality on picocyanobacterial abundance and community composition over a wider part of the study area but also to highlight potential effects of large multi-year oceanographic events, such as El Niño.

Experimental Procedures

Sample collection

Samples were collected from surface waters (5-10 m depth) of the western Pacific Ocean and its marginal seas in a total of nine cruises (Table S2, Fig. 1A). With the exception of the cruise in the Tokara Strait, which was conducted in November, all other cruises were performed in late spring to summer. Water was collected using Niskin bottles (12 L) attached to a conductivity, temperature, and depth (CTD) rosette multi-sampler (Sea Bird Electronics, USA). At each station, 0.5-3 L of seawater was pre-filtered through a 3.0 μm (47 mm) polycarbonate membrane (PALL Corporation) and then filtered onto a 0.22 μm (47 mm) polycarbonate membrane for DNA extraction. Membranes were frozen at -80 °C immediately after filtration. Temperature and salinity of seawater were measured using a conductivity-temperature-depth rosette system (CTD, Sea Bird Electronics). Seawater (1.8 mL) for flow cytometry analysis was fixed with 0.5% (final concentration) seawater-buffered paraformaldehyde. All samples were frozen in -80 °C until analysis. Inorganic nutrients including NO_3^-+NO_2^- (limit of detection (LOD): 0.1 μmol L^{-1}) and PO_4^{3-} (LOD: 0.08 μmol L^{-1}) were analyzed using the Technicon AA3 Auto-Analyzer (Bran+Luebbe, Germany) onboard or a QuAAtro Auto-Analyzer (Bran+Luebbe, Germany) on shore. Seawater samples for nutrient analysis were filtered through 0.45 μm acetate fiber membranes, except for the cruises in the Bering Sea, Tokara Strait and Sea of Okhotsk, for which samples were not prefiltered before measurements.

Analysis of *Synechococcus* abundance

Synechococcus cells were enumerated using a Becton-Dickinson FACSCalibur flow cytometer equipped with dual lasers of 488 nm and 635 nm with high flow rate (Liu *et al.*,

2014). Ten microliters of yellow-green fluorescent beads (1 μm, Polysciences, Warrington, PA, USA) were added to each sample as an internal standard. Flow cytometric data were analyzed using WinMDI software 2.9 (Joseph Trotter, Scripps Research Institute, La Jolla, CA, USA). All *Synechococcus* abundance data were used to generate a contour plot (Fig. 1B) using the contouring and shading of the Weighted-average gridding algorithm in Ocean Data View (Schlitzer, 2009). X and Y scale-lengths were set as 10 per-km. The relationship between the abundance of *Synechococcus* and environmental factors (Spearman's rank correlation coefficient) was analyzed by using SPSS (IBM SPSS Statistics Inc., Chicago). Data with total inorganic nitrogen (TIN) concentration lower than LOD (0.1 μmol L^{-1}) were set as 0.05 μmol L^{-1} and with PO_4^{3-} concentration lower than LOD (0.08 μmol L^{-1}) were set as 0.04 μmol L^{-1}.

Clone library construction and sequencing of *cpeBA* operon

DNA was extracted using the enzyme/phenol-chloroform protocol described elsewhere (Riemann *et al.*, 2000). Primer sets SynB3FW (5'-TCAAGGAGACCTACATCG-3') and SynA1R (5'-CAGTAGTTGATCAGRCGCAGGT-3') were used to amplify the *cpeBA* operon sequences (Everroad and Wood 2006). PCR reaction was carried out in 50 μL master mix including 5 μL of 10× buffer, 2 μL of $MgCl_2$ (25 mM), 4 μL of dNTPs (2.5 mM), 0.2 μl of Platinum Taq DNA polymerase (5 U, Invitrogen, USA), 1 μL of each primer (10 μM), 1 μL of DNA template (around 20 ng/μL) and 35.8 μL water. PCR products were purified using a PureLinkTM Quick Gel Extraction Kit (Invitrogen, USA). Purified DNA sequences were cloned into the PCR4.0 vector by using a TOPO TA cloning kit (Invitrogen, USA). In total, twelve clone libraries were constructed (Table S3). Forty to sixty positive clones from each

561 library were purified with a PureLink quick plasmid miniprep kit (Invitrogen, USA) and
562 sequenced in the MingBo sequencing company (Shanghai, China).
563
564 **Phylogenetic and diversity analysis of *cpeBA* operon**
565 All *cpeBA* operon sequences were aligned using DNAman (Woffelman, 2004) and
566 trimmed to equal length (506 bp). Chimeras were checked and removed using Mothur (Schloss
567 *et al.*, 2009; Edgar *et al.*, 2011). The operational taxonomic units (OTUs) numbers were
568 calculated at the cut-off level of 95% nucleotide identity. The representative sequence of each
569 OTU was randomly extracted and then identified by using BLAST searches against the
570 National Center for Biotechnology Information (NCBI) database
571 (http://www.ncbi.nlm.nih.gov). OTUs for which no significant similarity (E-value>10) was
572 found in NCBI database and heterotrophic bacteria sequences were removed. Abundance of
573 left OTUs was used to calculate Shannon index using Primer 5 (Primer-E-Ltd, UK). The
574 Shannon diversity index was calculated as follows: $H = -\sum Pi \ln Pi$, where Pi=S/N, S= number
575 of sequences of one OTU, N= total number of sequences in the sample (Shannon, 2001).
576 Representative OTU sequences were used for phylogenetic analyses. Maximum
577 Likelihood phylogenetic tree was constructed with MEGA 6 (Kumar *et al.*, 1994) using the
578 K2+G+I model for nucleotide evolution with 200 bootstraps. The most similar reference
579 sequences were retrieved from the NCBI database. The sequences were aligned using ClustalW
580 according to codon structure (Higgins *et al.*, 1994). Strain pigment information was derived
581 from published literature (Everroad and Wood, 2012; Humily *et al.*, 2013; Humily *et al.*, 2014)
582 and from the Roscoff Culture Collection (http://roscoff-culture-

583 collection.org/strains/shortlists/taxonomic-groups/marine-synechococcus).

584

585 **PCR and 454 sequencing**

586 Amplification of the *rpoC1* gene sequences was performed as previously described
587 (Mühling *et al.*, 2006). The first round of PCR used the primer *rpoC1*-N5 and the C-terminal
588 primer *rpoC1*-C, and the PCR products were used as templates for a second round of PCR with
589 primer *rpoC1*-39F (5'-adaptor A+barcode+GGNATNGTNTGYGAGCGYTG) and *rpoC1*-
590 462R (5'-adaptor B+CGYAGRCFCTTGRTCAGCTT (Mühling *et al.*, 2006; Xia *et al.*, 2015b).
591 PCR products were gel-purified using the Qiaquick gel purification kit, as described by the
592 manufacturer (Qiagen, Germany). Library quantification was done by fluorometry using the
593 Quant-iT picoGreen dsDNA Assay Kit (Invitrogen, USA). Amplicons were mixed in equal
594 amounts and sequenced in a two-region 454 run on a GS PicoTiterPlate using a GS Junior
595 pyrosequencing system according to manufacturer instructions (Roche, 454 Life Sciences,
596 Branford, CT, USA). The number of sequences obtained from each sample is listed in Table
597 S3.

598

599 **454 Post-run Sequence analysis**

600 Analysis of *rpoC1* sequences was conducted using the microbial ecology community
601 software program Mothur (Schloss *et al.*, 2009). Raw sequences were processed by removing
602 barcodes and primers only reads with an average quality score above 20 and read lengths
603 between 300 nt and 500 nt were taken into account. Sequence denoise was carried out using
604 the command shhh.seqs with sigma value of 0.01. Chimeras were analyzed using the command

chimera.uchime and removed. After the above quality control, sequences were identified by local Blast using BioEdit (Hall 1999) with an expectation value 1.0E-100. Reference sequences of each lineage are listed in Table S4. Sequences displaying less than 80% identity to the reference sequences or classified as *Prochlorococcus* and *Synechocystis* were removed. We then subsampled 1,739 reads from each sample for calculating DNA distance. A "shared" file was generated using Mothur's *make.shared* routine. OTU numbers were calculated at the cutoff level of 95% nucleotide identity. The Shannon diversity index was calculated as explained above, then after removing OTUs containing only one sequence, the Bray-Curtis similarity between samples was also computed. The Bray-Curtis similarity matrix was used to carry out NMDS analysis using Primer 5 (Primer-E-Ltd, UK).

In order to calculate the relative abundance of each phylogenetic lineage of *Synechococcus*, the representative sequence of each OTU was randomly extracted using command *get.oturep* and then identified by local Blast using Bio-Edit (Hall, 1999). The phylotype of each reference sequence is listed in Table S4. Since the similarity of sequences within each lineage ranged from 90% to 100% (Table S5), sequences displaying less than 90% identical to the reference sequences were assigned as unclassified (Xia *et al.*, 2015b). The relative abundance of each lineage was calculated and used to generate a heatmap using HemI (Deng *et al.*, 2014), then square root transformed in order to perform average linkage clustering using Pearson correlation matrices. Clade I *rpoC1* sequences were aligned according to codon structure and a Maximum Likelihood (ML) phylogenetic tree was constructed using MEGA 6 (Kumar *et al.*, 1994) with a K2P+G model. A heatmap showing the relative abundance of each OTU was generated using iTol (Letunic and Bork, 2007).

627

628 **Redundancy and correlation analysis**

629 The relationship between measured environmental parameters (Table S7) and
630 *Synechococcus* community composition was studied by redundancy analysis (RDA) using
631 CANOCO V4.5 (Microcomputer Power, USA). Stations that did not have nutrient data were
632 not included. Data with total inorganic nitrogen (TIN) concentration lower than LOD (0.1 μmol
633 L^{-1}) were set as 0.05 μmol L^{-1} and with PO_4^{3-} concentration lower than LOD (0.08 μmol L^{-1})
634 were set as 0.04 μmol L^{-1}. The matrix was generated using the relative abundance of each
635 lineage transformed by square root. Environmental data were normalized using z-score
636 transformation. The significance of the eigenvalues and species-environment correlations of
637 the first three axes were determined by Monte Carlo tests (500 permutations).

638 The spearman correlation between pigment types and phylogenetic lineages was analyzed
639 using the corrplot R-package (Wei, 2011). S5.2 and clade VIII which are known to mainly
640 consist of type 1 *Synechococcus* (Fuller *et al*., 2003; Dufresne *et al*., 2008) were not included
641 in the analysis.

642 **Sequences submission**

643 All sequences obtained from this study have been deposited in the National Center for
644 Biotechnology Information (NCBI) Sequence Read Archive (SRA) and Genbank (Table S6).

645

646

647

648

Acknowledgements:

We thank the captain and crew of the R/V Dongfanghong 2 (China), R/V Hakuho Maru (Japan), R/V Tansei Maru (Japan), and R/V Professor Multanovskiy (Russia) for offering opportunity for sampling during the cruises. We thank Dr. Dong Han Choi for providing pigment information of *Synechococcus* sp. strain KORDI-100 and Dr. Hisashi Endo providing nutrient data of Bearing Sea cruise. We also thank Candy Li for measuring *Synechococcus* cell abundance in Japan Sea, Eric for collecting DNA samples from Japan Sea and Wang Guo for sequencing *ropC1* gene sequences from samples collected from Tokara Strait.

This study was supported by the National Basic Research Program ("973" Program) of China through grant no. 2009CB421203 and the Research Grants Council of Hong Kong RGF grants (661912 and 661813) provided to H. Liu. F. Partensky and L. Garczarek were funded by the French "Agence Nationale de la Recherche" program BioAdapt SAMOSA (ANR-13-ADAP-0010). Also this work was partly supported by the Grant-in-Aid for #22221001, #22681004, and the project "The study of Kuroshio ecosystem dynamics for sustainable management" from MEXT, Japan.

675 **Conflict of interest**: The authors declare that there is no conflict of interest.

676

677 **References**

678 Agawin, N.S., Duarte, C.M., Agusti, S. (2000) Nutrient and temperature control of the
679 contribution of picoplankton to phytoplankton biomass and production. *Oceanogr* **45:**
680 591-600.

681 Ahlgren, N.A., Rocap, G. (2006) Culture isolation and culture-independent clone libraries
682 reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies.
683 *Appl Environ Microbiol* **72:** 7193-7204.

684 Ahlgren, N.A., Rocap, G. (2012) Diversity and distribution of marine *Synechococcus*: multiple
685 gene phylogenies for consensus classification and development of qPCR assays for
686 sensitive measurement of clades in the ocean. *Front Microbiol* **3**: 213.

687 Ahlgren, N.A., Noble, A., Patton, A.P., Roache-Johnson, K., Jackson, L., Robinson, D. *et al.*,
688 (2014) The unique trace metal and mixed layer conditions of the Costa Rica upwelling
689 dome support a distinct and dense community of *Synechococcus*. *Limnol Oceanogr* **59:**
690 2166-2184.

691 Buitenhuis, E.T., Li, W.K., Vaulot, D., Lomas, M.W., Landry, M., Partensky, F., *et al.,* (2012)
692 Picophytoplankton biomass distribution in the global ocean. *Earth Syst Sci Data* **4:** 37-46.

693 Campbell, L., and Iturriaga, R. (1988) Identification of Synechococcus spp. in the Sargasso
694 Sea by immunofluorescence and fluorescence excitation spectroscopy performed on
695 individual cells. *Limnol Oceanogr* 33: 1196-1201.

696 Chen, B., Liu, H., Landry, M.R., Dai, M., Huang, B., Sune, J. (2009) Close coupling between
697 phytoplankton growth and microzooplankton grazing in the western South China Sea.
698 *Limnol Oceanogr* **54:** 1084-1097.

699 Chen, F., Wang, K., Kan, J., Suzuki, M.T., Wommack, K.E. (2006) Diverse and unique
700 picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed
701 spacer sequences. *Appl Environ Microbiol* **72:** 2239-2243.

702 Choi, D.H., Noh, J.H. (2009) Phylogenetic diversity of *Synechococcus* strains isolated from
703 the East China Sea and the East Sea. *FEMS Microbiol Ecol* **69:** 439-448.

Chung, C-C., Huang, C-Y., Gong, G-C., Lin, Y-C. (2014) Influence of the Changjiang River flood on *Synechococcus* ecology in the surface waters of the East China Sea. *Microb Ecol* **67:** 273-285.

Chung, C-C., Gong, G-C., Huang, C-Y., Lin, J-Y., Lin, Y-C. (2015) Changes in the *Synechococcus* assemblage composition at the surface of the east China Sea due to flooding of the Changjiang river. *Microb Ecol***:** 1-12.

Cummins, P.F., and Freeland, H.J. (2007) Variability of the North Pacific Current and its bifurcation. *Prog Oceanogr* **75:** 253-265.

Deng, W., Wang, Y., Liu, Z., Cheng, H., Xue, Y. (2014) Hemi: A toolkit for illustrating heatmaps. *PloS One* **9:** e111988.

Dufresne, A, Ostrowski, M., Scanlan, D.J., Garczarek, L., Mazard, S., Palenik, B.P. *et al.*, (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9:** R90.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194-2200.

Everroad, C., Michelle, W.A. (2006) Comparative molecular evolution of newly discovered picocyanobacterial strains reveals a phylogenetically informative variable region of β-phycoerythrin. *J Phycol* **42:** 1300-1311.

Everroad, C., Six, C., Partensky, F., Thomas, J-C., Holtzendorff, J., Wood, A.M. (2006) Biochemical bases of type IV chromatic adaptation in marine *Synechococcus spp*. *J Bacteriol* **188:** 3345-3356.

Everroad, C., Wood, A.M. (2012) Phycoerythrin evolution and diversification of spectral phenotype in marine *Synechococcus* and related picocyanobacteria. *Mol Phylogenet Evol* **64:** 381-392.

Farrant G.K., Doré H., Cornejo-Castillo F.M., Partensky F., Ratin M., Ostrowski M. *et al.* (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci USA* **113**; E3365–E3374.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., *et al.,* (2013) Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceed Natl Acad Sci USA* **110:** 9824-9829.

734 Fuller, N.J., Marie, D., Partensky, F., Vaulot, D., Post, A.F., Scanlan, D.J. (2003) Clade-specific
735 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine
736 *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ*
737 *Microbiol* **69:** 2430-2443.
738 Fuller, N.J., Tarran, G.A., Yallop, M., Orcutt, K.M., Scanlan, D.J. (2006) Molecular analysis
739 of picocyanobacterial community structure along an Arabian Sea transect reveals distinct
740 spatial separation of lineages. *Limnol Oceanogr* **51:** 2515-2526.
741 Garcia-Pichel, F., Belnap, J., Neuer, S., Schanz, F. (2003) Estimates of global cyanobacterial
742 biomass and its distribution. *Algolog Stud* **109:** 213-227.
743 Gascuel, O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model
744 of sequence data. *Mol Biol Evol* **14:** 685-695.
745 Guo, C., Liu, H., Zheng, L., Song, S., Chen, B., and Huang, B. (2014) Seasonal and spatial
746 patterns of picophytoplankton growth, grazing and distribution in the East China Sea.
747 *Biogeosciences* 11: 1847-1862.
748 Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis
749 program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41:** 95-98.
750 Haverkamp, T., Acinas, S.G., Doeleman, M., Stomp, M., Huisman, J., Stal, L.J. (2008)
751 Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and
752 phycobiliprotein operons. *Environ Microbiol* **10:** 174-188.
753 Haverkamp, T.H., Schouten, D., Doeleman, M., Wollenzien, U., Huisman, J., Stal, L.J. (2009)
754 Colorful microdiversity of *Synechococcus* strains. (picocyanobacteria) isolated from the
755 Baltic Sea. *ISME J* **3:** 397-408.
756 Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G,, Gibson, T.J. (1994)
757 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment
758 through sequence weighting, position-specific gap penalties and weight matrix choice.
759 *Nucleic Acids Res* **22:**4673-4680.
760 Huang, S., Wilhelm, S.W., Harvey, H.R., Taylor, K., Jiao, N., Chen, F. (2012) Novel lineages
761 of *Prochlorococcus* and *Synechococcus* in the global oceans. *ISME J* **6:** 285-297.
762 Humily, F., Partensky, F., Six, C., Farrant, G.K., Ratin, M., Marie, D., *et al.,* (2013) A gene
763 island with two possible configurations is involved in chromatic acclimation in marine

764 *Synechococcus*. *PloS One* **8**: e84459.
765 Humily, F., Farrant, G.K., Marie, D., Partensky, F., Mazard, S., Perennou, M., *et al.,* (2014)
766 Development of a targeted metagenomic approach to study a genomic region involved in
767 light harvesting in marine *Synechococcus*. *FEMS Microbiol Ecol* **88:** 231-249.
768 Hunter-Cevera, K.R., Post, A.F., Peacock, E.E., Sosik, H.M. (2015) Diversity of
769 *Synechococcus* at the Martha's Vineyard coastal observatory: Insights from culture
770 isolations, clone libraries, and flow cytometry. *Microb Ecol***:** 1-14.
771 Jiao, N., Yang, Y., Hong, N., Ma, Y., Harada, S., Koshikawa, H., *et al.,* (2005) Dynamics of
772 autotrophic picoplankton and heterotrophic bacteria in the East China Sea. *Cont Shelf Res*
773 **25:** 1265-1279.
774 Kirk. J.T. (1994) *Light and photosynthesis in aquatic ecosystems*. Cambridge university press.
775 509 pp.
776 Kumar, S., Tamura, K., Nei, M. (1994) MEGA: molecular evolutionary genetics analysis
777 software for microcomputers. *Computer applications in the biosciences: CABIOS* **10:**
778 189-191.
779 Lantoine, F., Neveux, J. (1997) Spatial and seasonal variations in abundance and spectral
780 characteristics of phycoerythrins in the tropical northeastern Atlantic Ocean. *Deep Sea*
781 *Res Pt I* **44:** 223-246.
782 Larsson, J., Celepli, N., Ininbergs, K., Dupont, C.L., Yooseph, S., Bergman, B., *et al.,* (2014)
783 Picocyanobacteria containing a novel pigment gene cluster dominate the brackish water
784 Baltic Sea. *ISME J* **8:** 1892-1903.
785 Letunic, I., Bork, P. (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic
786 tree display and annotation. *Bioinformatics* **23:** 127-128.
787 Li, Q., Legendre, L., Jiao, N. (2015) Phytoplankton responses to nitrogen and iron limitation
788 in the tropical and subtropical Pacific Ocean. *J Plankt Res* **37:** 306-319.
789 Li, WK. (1998) Annual average abundance of heterotrophic bacteria and *Synechococcus* in
790 surface ocean waters. *Limnol Oceanogr* **43:** 1746-1753.
791 Liu, H., Campbell, L., Landry, M., Nolla, H., Brown, S., Constantinou, J. (1998)
792 *Prochlorococcus* and *Synechococcus* growth rates and contributions to production in the
793 Arabian Sea during the 1995 Southwest and Northeast Monsoons. *Deep Sea Res Pt II* **45:**

794 2327-2352.
795 Liu, H., Jing, H., Wong, T.H., Chen, B. (2014) Co-occurrence of phycocyanin-and
796 phycoerythrin-rich *Synechococcus* in subtropical estuarine and coastal waters of Hong
797 Kong. *Environ Microbiol Rep* **6:** 90-99.
798 Mazard, S., Ostrowski, M., Partensky, F., Scanlan, D.J. (2012) Multi-locus sequence analysis,
799 taxonomic resolution and biogeography of marine *Synechococcus*. *Environ Microbiol* **14:**
800 372-386.
801 Mella-Flores, D., Mazard, S., Humily, F., Partensky, F., Mahé, F., Bariat, L., *et al.,* (2011) Is
802 the distribution of *Prochlorococcus* and *Synechococcus* ecotypes in the Mediterranean Sea
803 affected by global warming? *Biogeosciences* **8:** 2785-2804.
804 Mühling, M., Fuller, N.J., Somerfield, P.J., Post, A.F., Wilson, W.H., Scanlan, D.J., *et al.,*
805 (2006) High resolution genetic diversity studies of marine *Synechococcus* isolates using
806 rpoC1-based restriction fragment length polymorphism. *Aquat Microb Ecol* **45:** 263-275.
807 Nagata, Y., Takeshita, K. (1985) Variation of the sea surface temperature distribution across the
808 Kuroshio in the Tokara Strait. *J Oceanogr Soc Japan* **41:** 244-258.
809 Nosaka, Y., Isada, T., Kudo, I., Saito, H., Hattori, H., Tsuda, A., *et al.,* (2014) Light utilization
810 efficiency of phytoplankton in the Western Subarctic Gyre of the North Pacific during
811 summer. *J oceanogr* **70:** 91-103.
812 Olson, R., Chisholm, S., Zettler, E., Armbrust, E. (1988) Analysis of *Synechococcus* pigment
813 types in the sea using single and dual beam flow cytometry. *Deep Sea Research* **35:** 425-
814 440.
815 Ong, L., and Glazer, A. (1991) Phycoerythrins of marine unicellular cyanobacteria. I. Bilin
816 types and locations and energy transfer pathways in *Synechococcus* spp. phycoerythrins.
817 *J Biol Chem* 266: 9515-9527.
818 Partensky, F., Blanchot, J., Lantoine, F., Neveux, J., Marie, D. (1996) Vertical structure of
819 picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean.
820 *Deep Sea Res Pt I* **43:** 1191-1213.
821 Partensky, F., Blanchot, J., Vaulot, D. (1999) Differential distribution and ecology of
822 *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Bull Inst Oceanogr*
823 *Monaco* **19:** 457-476.

824 Penno, S., Lindell, D., Post, A.F. (2006) Diversity of *Synechococcus* and *Prochlorococcus*
825 populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ*
826 *microbiol* **8:** 1200-1211.
827 Pittera, J., Humily, F., Thorel, M., Grulois, D., Garczarek, L., Six, C. (2014) Connecting
828 thermal physiology and latitudinal niche partitioning in marine *Synechococcus*. *ISME J* **8:**
829 1221-1236.
830 Post, A.F., Penno, S., Zandbank, K., Paytan, A., Huse, S.M., Welch, D.M. (2011) Long term
831 seasonal dynamics of *Synechococcus* population structure in the Gulf of Aqaba, Northern
832 Red Sea. *Front Microbiol* **2:** 131.
833 Qiu, B. (2001) Kuroshio and Oyashio currents: Academic Press.
834 Riemann, L., Steward, G.F., Azam, F. (2000) Dynamics of bacterial community composition
835 and activity during a mesocosm diatom bloom. *Applied and Environ Microbiol* **66:** 578-
836 587.
837 Saito, M.A., Rocap, G., Moffett, J.W. (2005) Production of cobalt binding ligands in a
838 *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol Oceanogr* **50:** 279-290.
839 Sawada, K., and Handa, N. (1998) Variability of the path of the Kuroshio ocean current over
840 the past 25,000 years. *Nature* **392**: 592-595.
841 Scanlan, D.J., Ostrowski, M., Mazard, S., Dufresne, A., Garczarek, L., Hess, W.R., *et al.,*
842 (2009) Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73:**
843 Schlitzer, R. (2009) Ocean data view. Available at https://odv.awi.de/
844 Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al.,* (2009)
845 Introducing mothur: open-source, platform-independent, community-supported software
846 for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537-
847 7541.
848 Shannon, C.E. (2001) A mathematical theory of communication. *ACM SIGMOBILE Mob*
849 *Comput Commun Rev* 5: 3-55.
850 Six, C., Thomas, J.C., Garczarek, L., Ostrowski, M., Dufresne, A., Blot, N., *et al.,* (2007)
851 Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a comparative
852 genomics study. *Genome Biol* **8:** R259.
853 Sohm, J.A., Ahlgren, N.A., Thomson, Z.J., Williams, C., Moffett, J.W., Saito, M.A., *et al.,*

(2015) Co-occurring *Synechococcus* ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* **10:** 333-345.

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

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

Sudek, S., Everroad, R.C., Gehman, ALM., Smith, J.M., Poirier, C.L., Chavez, F.P., *et al.,* (2015) Cyanobacterial distributions along a physico-chemical gradient in the Northeastern Pacific Ocean. *Environ Microbiol* **17:** 3692-3707.

Tai, V., Palenik, B. (2009) Temporal variation of *Synechococcus* clades at a coastal Pacific Ocean monitoring site. *ISME J* **3:** 903-915.

Toledo, G., Palenik, B. (1997) *Synechococcus* diversity in the California current as seen by RNA polymerase (*rpoC1*) gene sequences of isolated strains. *Appl Environ Microbiol* **63:** 4298-4303.

Urbach, E., Scanlan, D.J., Distel, D.L., Waterbury, J.B., and Chisholm, S.W. (1998) Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *J Mol Evol* 46: 188-201.

Wang, K., Wommack, K.E., Chen, F. (2011) Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. *Appl and environ microbio* **77:** 7459-7468.

Wei T. (2011). Corrplot: visualization of a correlation matrix. R package version 0.77. https://CRAN.R-project.org/package=corrplot.

Woffelman, C. (2004) DNAMAN for Windows, Version 5.2. 10. *Lynon Biosoft, Institute of Molecular Plant Sciences, Netherlands: Leiden University*.

Wood, A.M. (1985) Adaptation of photosynthetic apparatus of marine ultraphytoplankton to natural light fields. *Nature* **316**: 253-255.

Wood, A.M., Phinney, D.A., and Yentsch, C.S. (1998) Water column transparency and the distribution of spectrally distinct forms of phycoerythrin-containing organisms. *Mar Ecol*

Prog Ser **162**: 25-31.

Wood, A.M., Lipsen, M., and Coble, P. (1999) Fluorescence-based characterization of phycoerythrin-containing cyanobacterial communities in the Arabian Sea during the Northeast and early Southwest Monsoon (1994–1995). *Deep Sea Res Pt II* **46**: 1769-1790.

Xia, X., Guo, W., and Liu, H. (2015a) Dynamics of the bacterial and archaeal communities in the Northern South China Sea revealed by 454 pyrosequencing of the 16S rRNA gene. *Deep Sea Res Pt II* **117**: 97-107.

Xia, X., Vidyarathna, N.K., Palenik, B., Lee, P., Liu H. (2015b) Comparison of the seasonal variation of *Synechococcus* assemblage structure in estuarine waters and coastal waters of Hong Kong. *Appl Environ Microbiol* **81:** 7644-7655.

Zwirglmaier, K., Jardillier, L., Ostrowski, M., Mazard, S., Garczarek, L., Vaulot, D., *et al.,* (2008) Global phylogeography of marine *Synechococcus* and *Prochlorococcus r*eveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10:** 147-161.

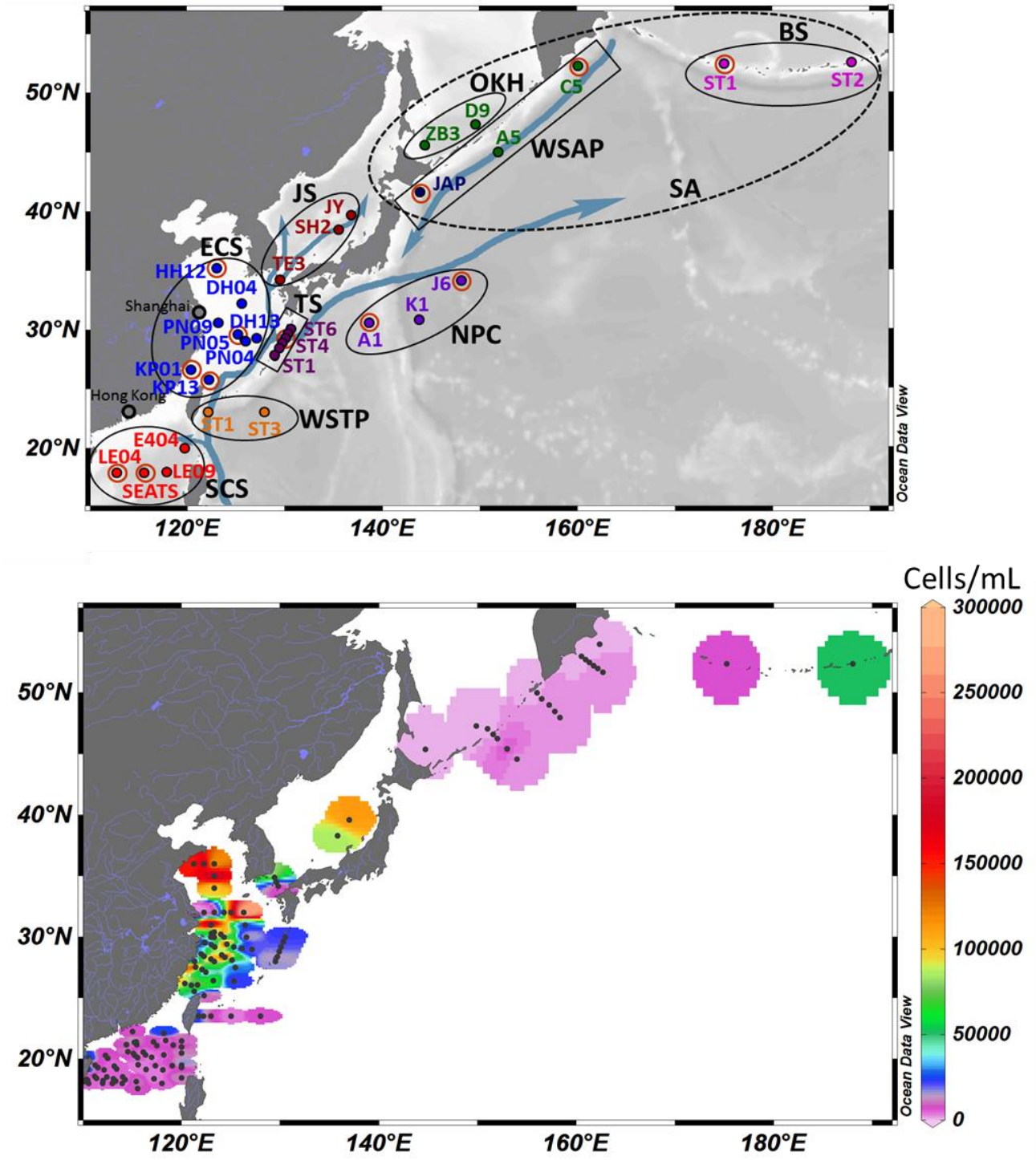
914 Fig. 1 Localization of studied sites in the northwestern Pacific Ocean. A, Samples used for
915 pyrosequencing of the *rpoC1* gene are shown as dots and those used for constructing *cpeBA*
916 operon clone libraries as circles. Samples from different cruises were labeled in different colors.
917 B, Contour plot (created with Ocean Data View (Schlitzer, 2009)) showing *Synechococcus* cell
918 concentrations in the northwestern Pacific Ocean using Weighted-average gridding over all
919 sampling stations. Black dots indicate sampling stations. South China Sea (SCS); East China
920 Sea (ECS); western subtropical Pacific Ocean (WSTP); Tokara Strait (TS); Japan Sea/East Sea
921 (JS); Northern Pacific Current (NPC); Sea of Okhotsk (OKH); western subarctic Pacific Ocean
922 (WSAP); Bering Sea (BS); subarctic ocean (SA) including BS, OKH and WSAP.
923
924 Fig. 2 Maximum-likelihood phylogenetic tree of *cpeBA* operon sequences of *Synechococcus*
925 obtained from the western Pacific Ocean. Representative sequences of 95% OTUs containing
926 more than 2 sequences were included in the dataset. Right bars show three clades formed by
927 *cpeBA* operon sequences obtained from the western Pacific Ocean. Strain pigment information
928 derived from Everroad *et al.,* (2012) and Humily *et al.,* (2014). Bootstrap values greater than
929 50% were shown on nodes of branches. Clade names for reference strains are given after strain
930 names.
931
932 Fig. 3 Distribution of *Synechococcus* pigment types in surface waters of the northwestern
933 Pacific Ocean, as indirectly assessed using the diversity of the *cpeBA* operon (see Fig. 2).
934 Pigment type assignment was made following the nomenclature proposed by Humily *et al.,*
935 (2014).
936
937 Fig. 4 NMDS plots showing the similarity of *Synechococcus* communities based on *rpoC1*
938 gene. Samples from different cruises were labeled in different colors. A, similarity of
939 *Synechococcus* communities in the western Pacific Ocean. B, similarity of *Synechococcus*
940 community in the tropic/subtropics oceans of the western Pacific Ocean.
941
942 Fig. 5 Heatmap displaying the relative abundances of *Synechococcus* lineages for any given
943 station of the northwestern Pacific Ocean, as assigned based on *rpoC1* gene. Data were

transformed by square root transformation. Unclassified sequences were not included.

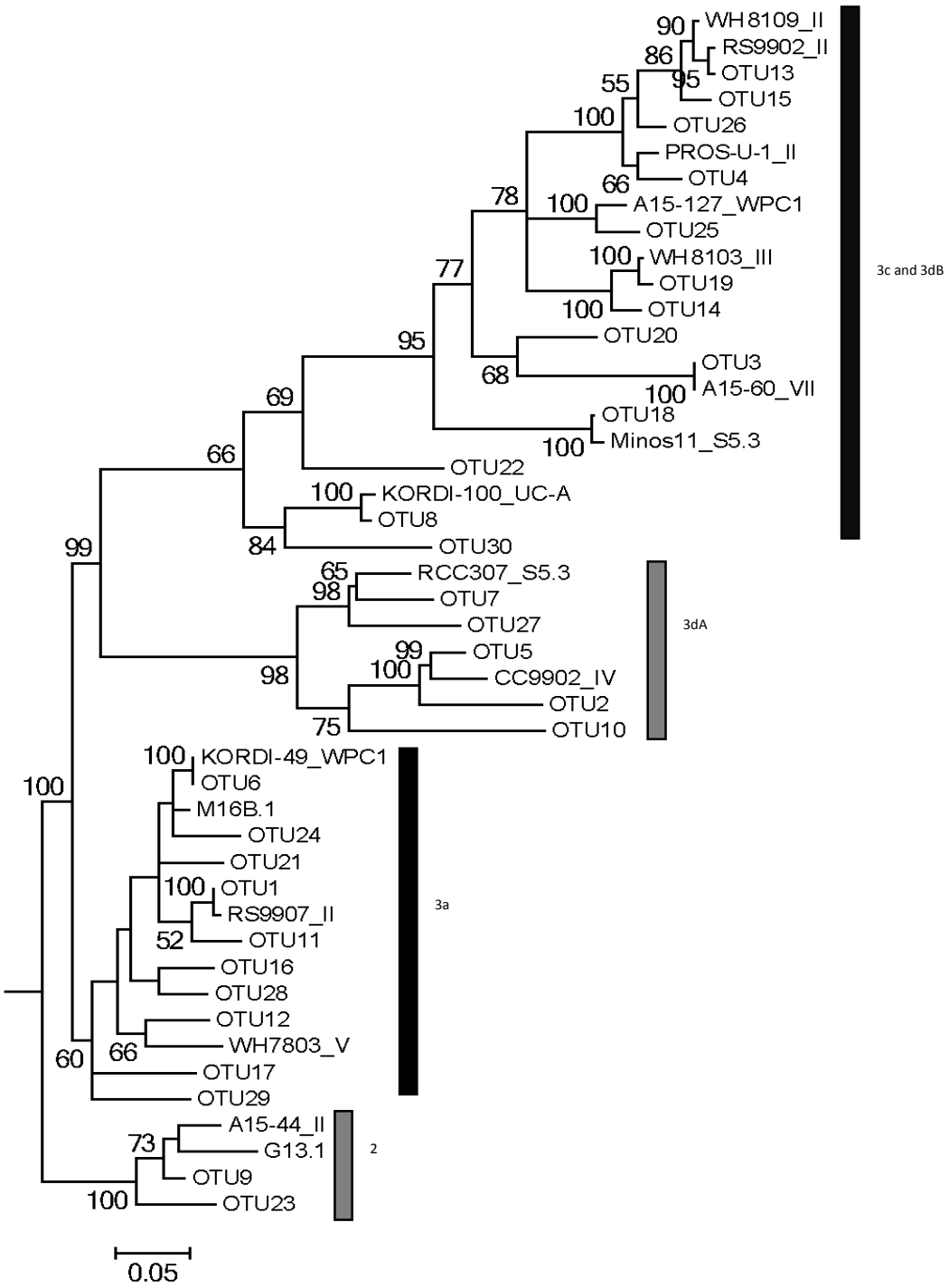
Fig. 6 Correlation biplots based on a redundancy analysis (RDA) depicting the relationship between the environmental factors and *Synechococcus* lineages. A. All samples with environmental data were analyzed. B. Samples with temperature higher than 25 ℃ were analyzed. Relative abundance of each lineage was normalized by square root-transformation. The environmental data were z-score transformed. *P<0.05, **P<0.01.

Fig. 7 Relative abundance of the different lineages versus latitude. A, Relative abundance of the two dominant clades (I and II). B, Relative abundance of all other major lineages.

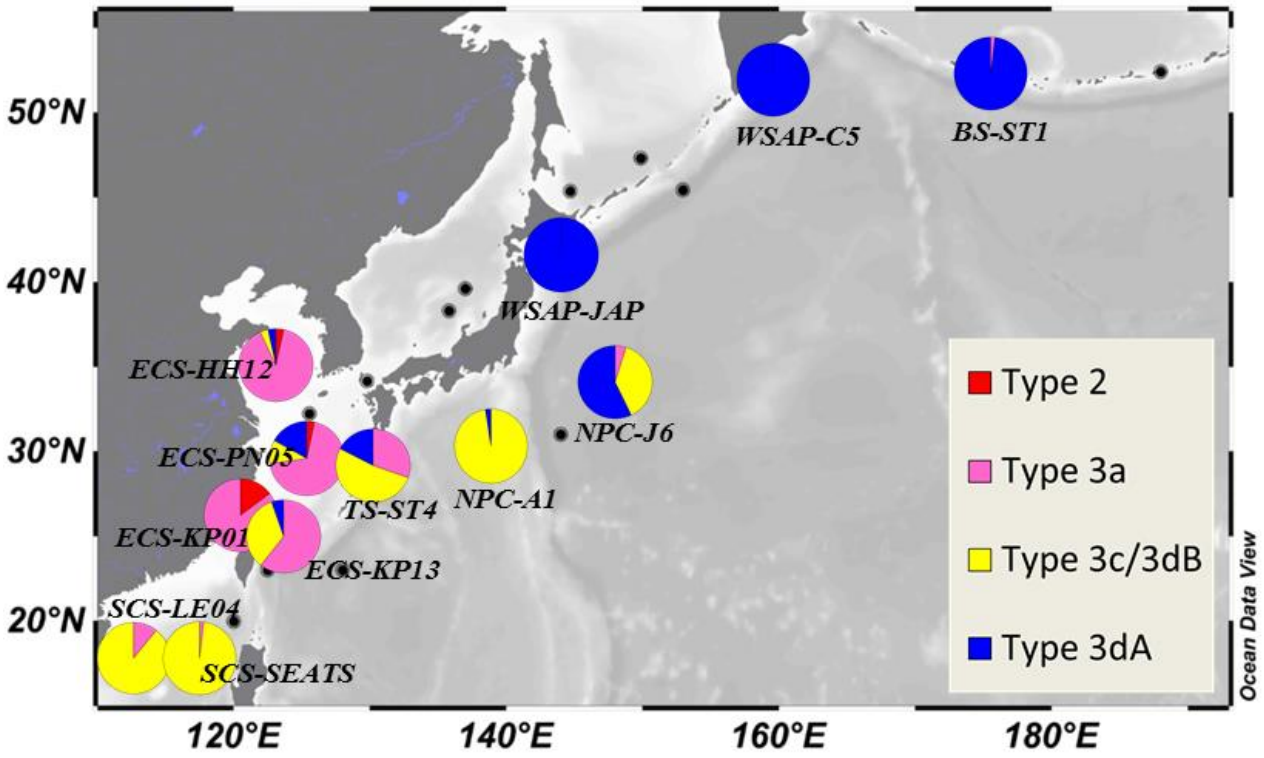
Fig. 8 Maximum-likelihood phylogenetic tree of clade I *Synechococcus* using the 100 most abundant *rpoC1* OTUs across all samples. Heatmap on right-hand side shows the relative abundance of OTUs in each library (square root transformed). Only nodes with bootstrap values higher than 50% are shown. Letters A-F on right-hand side correspond to different subclades within clade I.



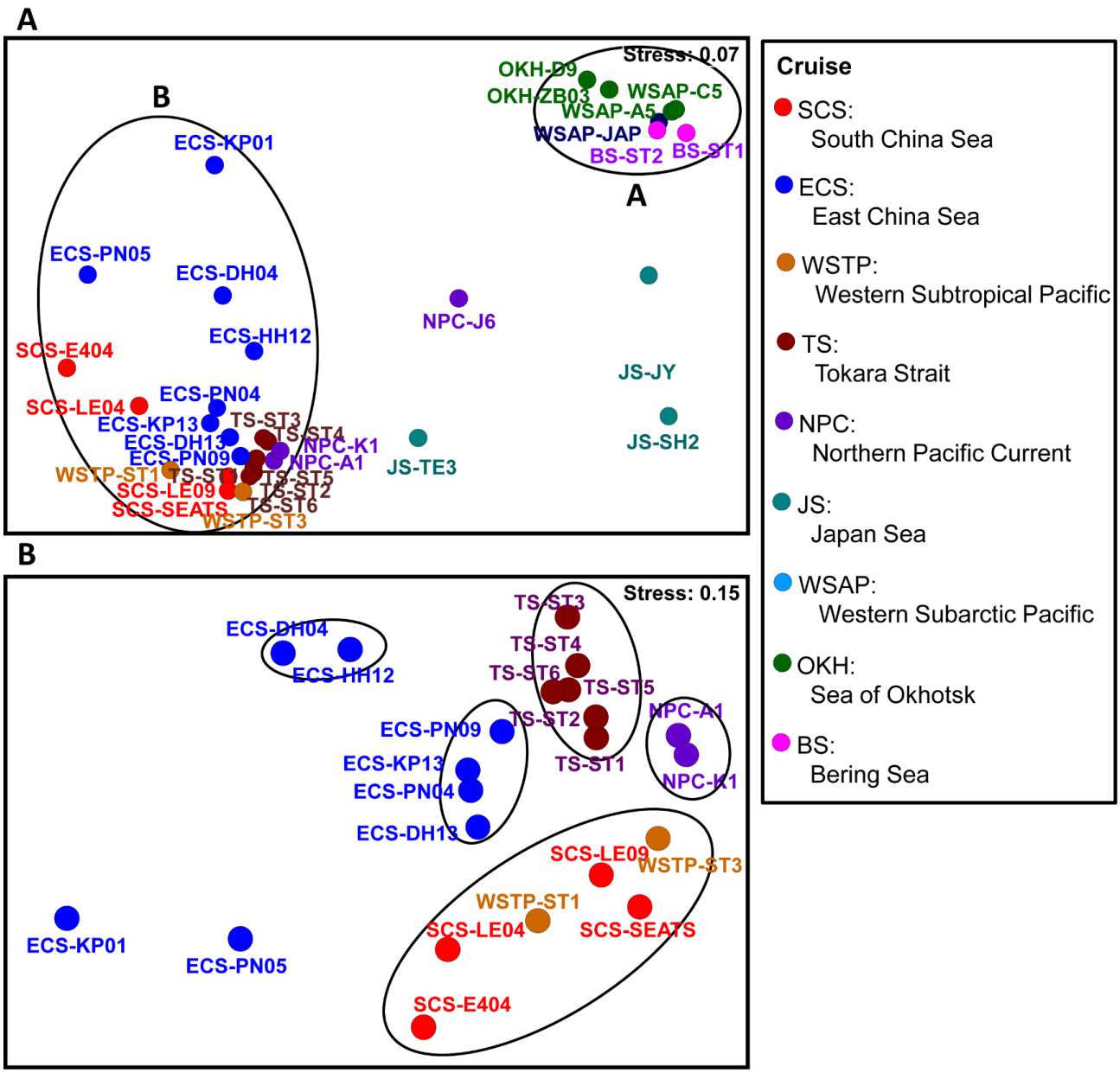
Xia et al. Fig. 1



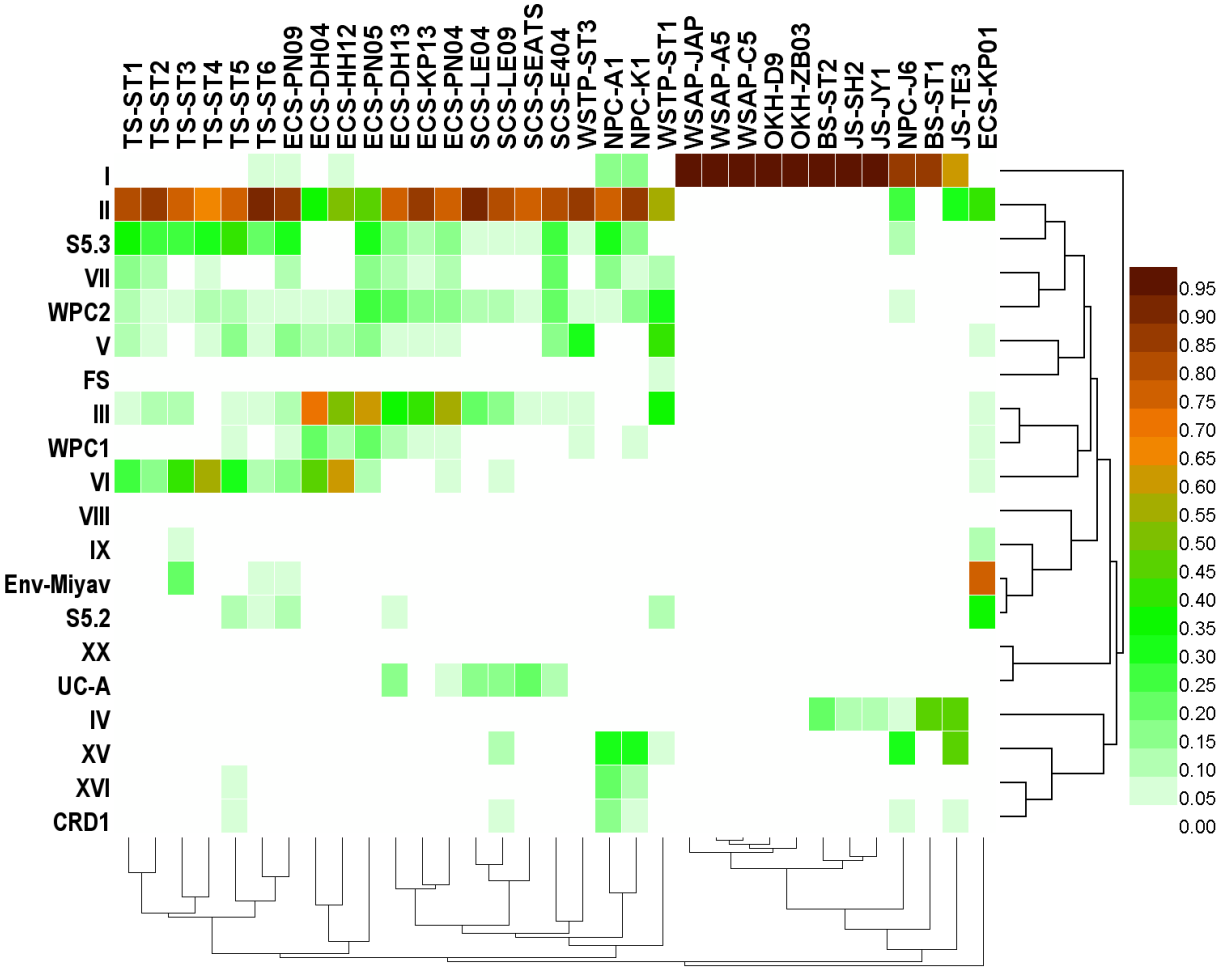
Xia et al. Fig. 2



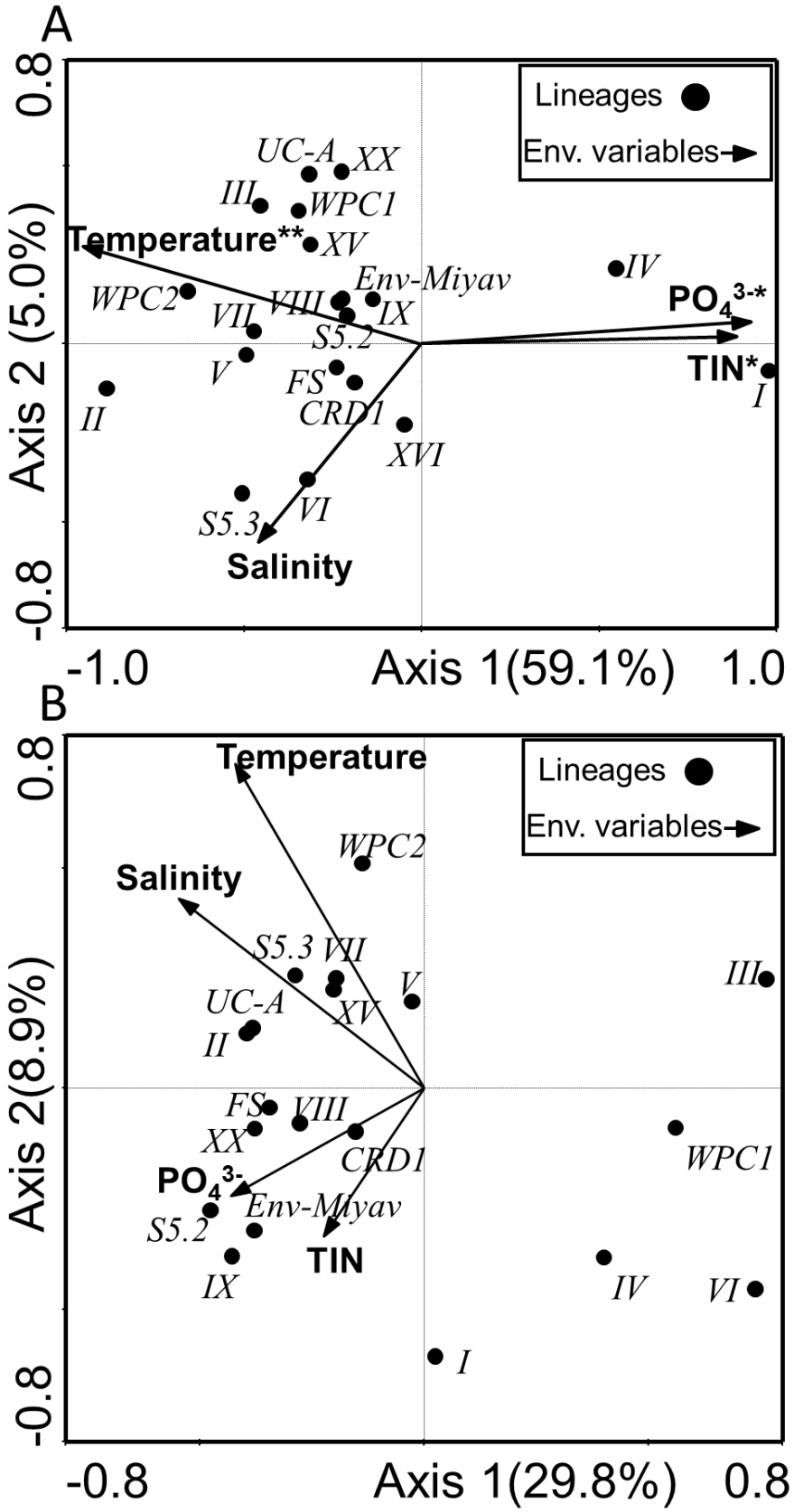
Xia et al. Fig. 3



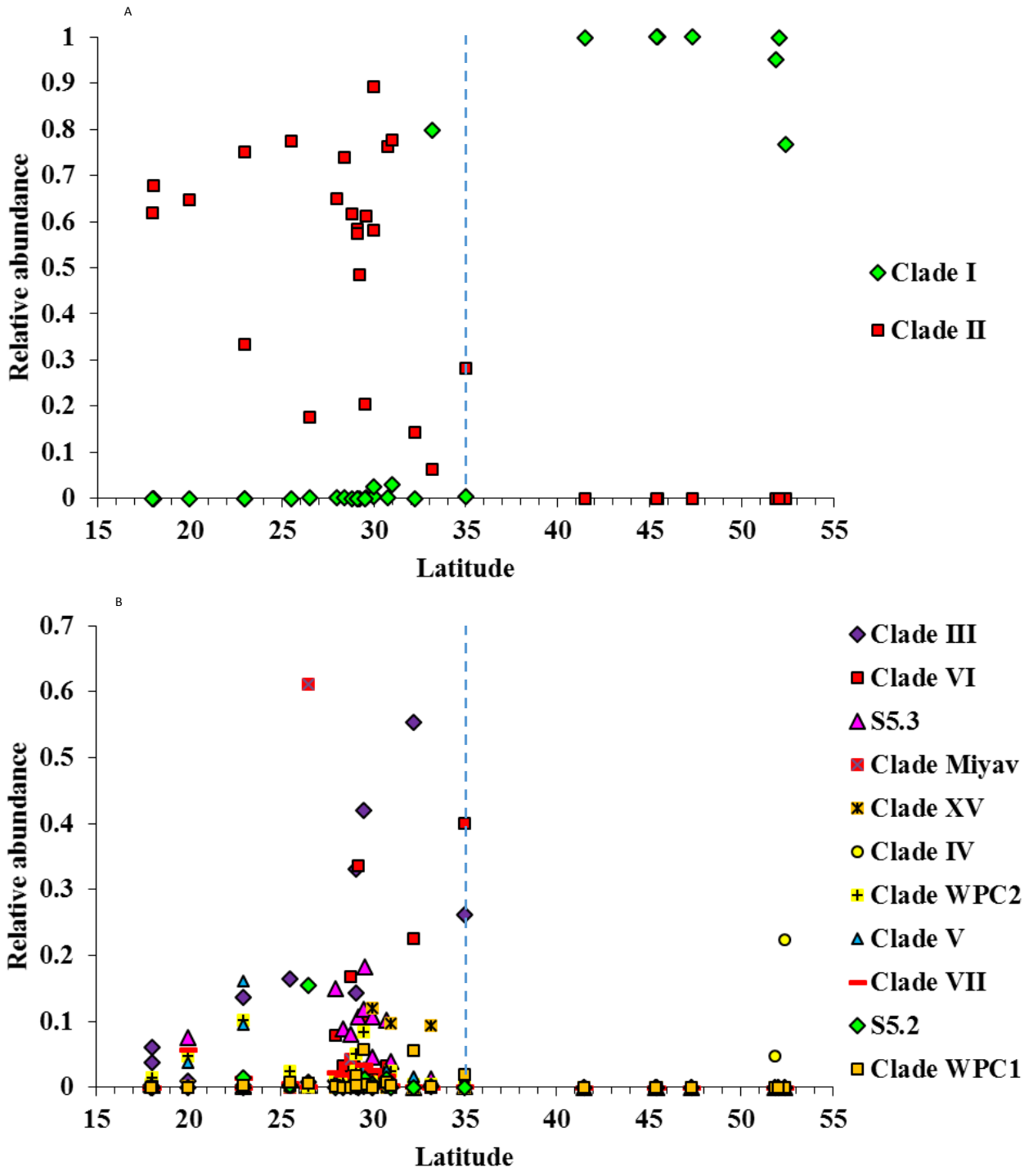
Xia et al. Fig. 4



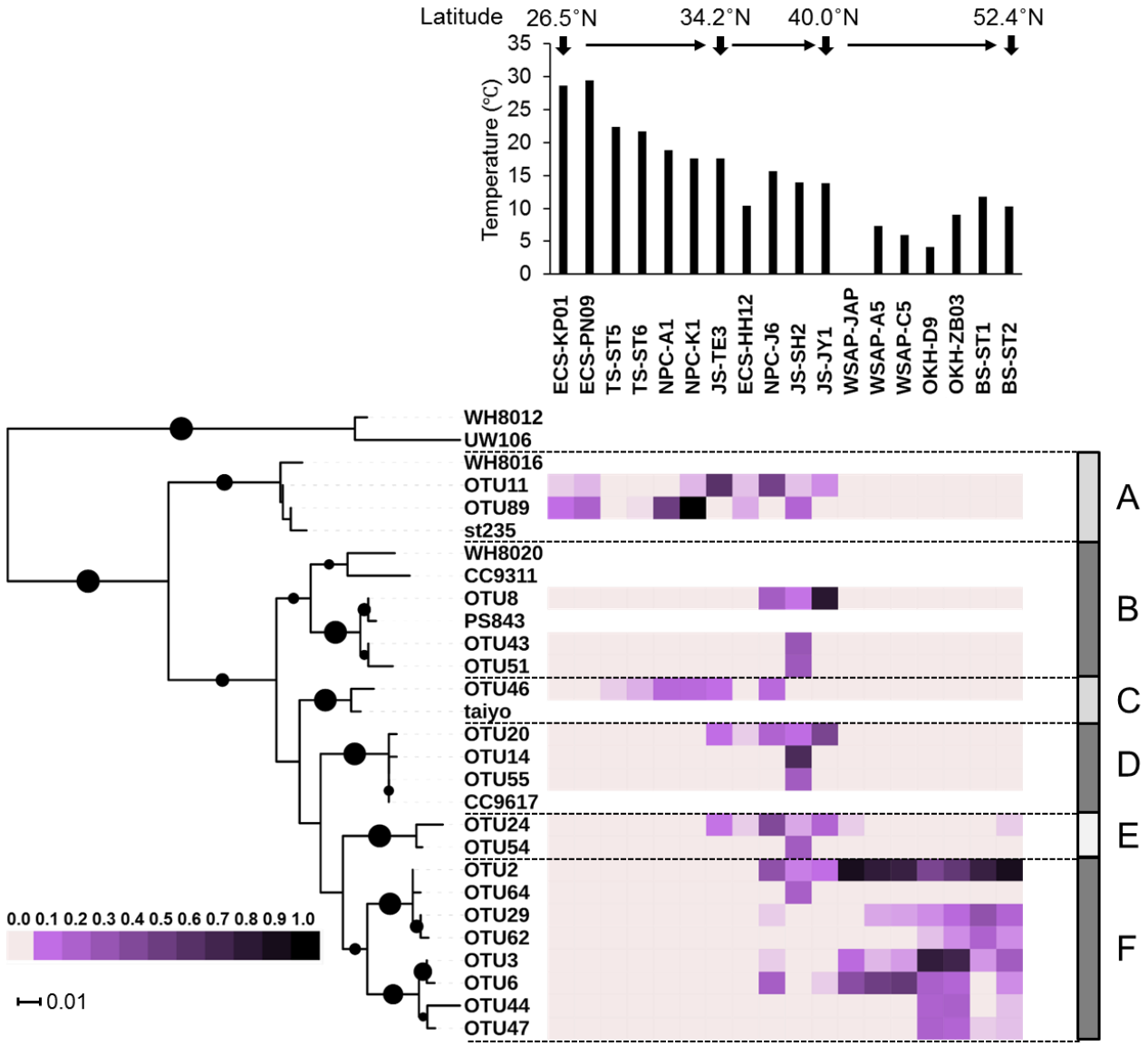
Xia et al. Fig. 5



Xia et al. Fig. 6



Xia et al. Fig. 7



Xia et al. Fig. 8