Environmental Microbiology (2008) **10**(1), 147–161 doi:10.1111/j.1462-2920.2007.01440.x

Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes

Katrin Zwirglmaier,[1] Ludwig Jardillier,[1] Martin Ostrowski,[1] Sophie Mazard,[1] Laurence Garczarek,[2] Daniel Vaulot,[2] Fabrice Not,[2,3] Ramon Massana,[3] Osvaldo Ulloa[4] and Dave J. Scanlan[1]*

[1]*Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK.*
[2]*Station Biologique de Roscoff, UMR 7144 CNRS and Université Pierre et Marie Curie Place Georges Teissier, BP74, 29682 Roscoff cedex, France.*
[3]*Department of Marine Biology and Oceanography, Institut de Ciències del Mar, CSIC, Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain.*
[4]*Department of Oceanography and Center for Oceanographic Research COPAS, University of Concepción, PO 160-C, Concepción, Chile.*

Summary

Marine cyanobacteria of the genera *Prochlorococcus* and *Synechococcus* are important contributors to global primary production occupying a key position at the base of marine food webs. The genetically diverse nature of each genus is likely an important reason for their successful colonization of vast tracts of the world's oceans, a feature that has led to detailed analysis of the distribution of these genetic lineages at the local and ocean basin scale. Here, we extend these analyses to the global dimension, using new data from cruises in the Pacific, Indian and Arctic Oceans in combination with data from previous studies in the Atlantic Ocean, Arabian Sea, Red Sea and a circumnavigation of the southern hemisphere to form a data set which comprises most of the world's major ocean systems. We show that the distribution patterns of *Prochlorococcus* and *Synechococcus* lineages are remarkably similar in different ocean systems with comparable environmental conditions, but producing a strikingly different 'signature' in the four major ocean domains or biomes (the Polar Domain, Coastal Boundary Domain, Trade Winds Domain and Westerly Winds Domain). This clearly reiterates the idea of spatial partitioning of individual cyanobacterial lineages, but at the global scale.

Received 21 May, 2007; accepted 14 August, 2007. *For correspondence. E-mail d.j.scanlan@warwick.ac.uk; Tel. (+44) 24 7652 8363; Fax (+44) 24 7652 3701.

Introduction

With cell numbers of up to 10^6 and 10^5 cells ml^{-1}, respectively, the unicellular cyanobacteria *Prochlorococcus* and *Synechococcus* are the most abundant representatives of picophytoplankton in marine systems responsible for up to 50% of CO_2 fixation in some oceanic regions (Li, 1994; Liu *et al*., 1997; Veldhuis *et al*., 1997).

Prochlorococcus was first isolated less than 20 years ago (Chisholm *et al*., 1988; 1992) and is probably the most abundant photosynthetic organism on Earth (for reviews see Partensky *et al*., 1999a; Scanlan and West, 2002). At least two distinct ecotypes are known, which can be traced by 16S rRNA gene phylogeny and differ markedly in their photophysiology and consequently pigment configuration (Moore and Chisholm, 1999; Partensky *et al*., 1999a). These ecotypes, termed high light (HL) and low light (LL) adapted, are further subdivided into HLI and HLII, and LLI, LLII, LLIII, LLIV (Rocap *et al*., 2002) or, more recently, into ecotype 'type strains' (Ahlgren *et al*., 2006b).

Synechococcus comprising marine subcluster 5.1 (Herdman *et al*., 2001), the most abundant *Synechococcus* throughout the euphotic zone in marine systems, are phylogenetically closely related to *Prochlorococcus*, together forming a discrete picophytoplankton clade (Urbach *et al*., 1998). The two genera are distinguishable however, by their possession of dissimilar light-harvesting apparatus (Ting *et al*., 2002). Based on 16S rRNA gene analysis, 10 lineages within the *Synechococcus* 5.1 subcluster have been described (Fuller *et al*., 2003), while five more potential lineages have been designated based on ITS and *ntcA* gene sequence analysis (Ahlgren and Rocap, 2006a; Penno *et al*., 2006) and this number appears likely to grow further (e.g. see Muhling *et al*., 2006). With this large number of phylogenetically distinct lineages, the community structure of *Synechococcus* is likely much more complex than that of *Prochlorococcus* and is currently still poorly understood.

Although certain aspects of the ecology of *Prochlorococcus* and *Synechococcus* are well documented, namely that (i) where they co-occur *Prochlorococcus* peak cell abundances are about one order of magnitude higher than *Synechococcus*, (ii) *Prochlorococcus* is more abundant in warm (> 15°C) oligotrophic waters, whereas *Synechococcus* dominates in coastal and more temperate/mesotrophic open ocean waters (see Partensky *et al*., 1999b) and (iii) data from several North-South Atlantic transects have shown that standing stocks of *Synechococcus* and *Prochlorococcus* are relatively stable and undergo little interannual variation (Heywood *et al*., 2006), it is only relatively recently that specific spatial distribution patterns of marine picocyanobacterial lineages have been reported (Fuller *et al*., 2005; 2006; Bouman *et al*., 2006; Johnson *et al*., 2006; Garczarek *et al*., 2007; Zwirglmaier *et al*., 2007). For *Prochlorococcus*, this has highlighted that specific vertical partitioning of HL and LL ecotypes occurs over large spatial scales in stratified water columns, and that the relative abundance of HL ecotypes is related to the structure and strength of this vertical mixing as well as with temperature. For *Synechococcus*, although reports of depth preferences of individual clades in specific oceanic regions exist (Ferris and Palenik, 1998; Toledo and Palenik, 2003), more clear is that horizontal partitioning of specific lineages occurs, as demonstrated from transect work conducted in both the Arabian Sea and Atlantic Ocean (Fuller *et al*., 2006; Zwirglmaier *et al*., 2007).

Understanding the genetic structure of these cyanobacterial picophytoplankton *in situ* is important for accurately modelling growth and carbon fixation by these organisms (Follows *et al*., 2007) but also for elucidating the biological and physical factors controlling the presence or absence of particular lineages in any particular locality. This latter goal overlaps with issues of microbial biogeography, particularly the endemic versus cosmopolitan nature of bacterial populations, a topic that is receiving increasing interest (see Dolan, 2005; Martiny *et al*., 2006). Furthermore, the recent availability of full genome sequences for several *Prochlorococcus* and *Synechococcus* strains (Dufresne *et al*., 2003; Rocap *et al*., 2003; Palenik *et al*., 2003; 2006) which, although already leading to new insights into mechanisms of niche adaptation (Johnson *et al*., 2006), still requires a more detailed ecological framework to fully facilitate the evolutionary context and bioinformatic 'mining' of such genetic data.

With these requirements in mind, we present here new spatial partitioning data for picocyanobacterial lineages in three oceanic systems, and then extend our analysis to the global scale by combining these three new data sets with previously published ones. This allowed unprecedented insight into the phylogeography of *Synechococcus* and *Prochlorococcus*. Statistical approaches were used to relate the distribution patterns to environmental gradients and various factors influencing the distribution of individual clades are discussed.

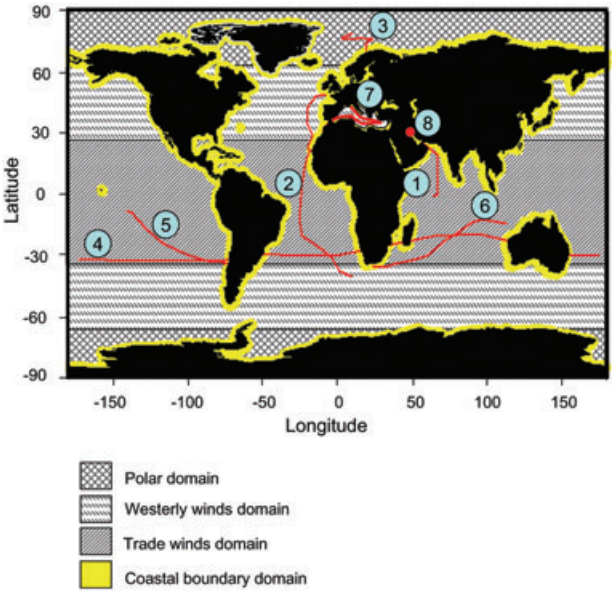

Fig. 1. Map of the cruise tracks analysed in this study. See Table 1 for further information. The four major ocean biomes as defined by Longhurst (1995; 2007) are also indicated.

Results and discussion

Geographic areas

We analysed samples from seven cruises that took place between 1999 and 2004 and one station in the northern Red Sea that was sampled at monthly intervals through an annual cycle (1999–2000). The cruise tracks are depicted in Fig. 1 with additional information summarized in Table 1. The geographic areas studied represent a number of the world's major ocean systems, including the Indian, eastern South Pacific, Atlantic and Arctic Oceans, covering the four major ocean domains (biomes) defined by Longhurst (1995; 2007). They also span different temperature zones ranging from arctic to temperate, subtropical and tropical, as well as contrasting oceanic provinces (defined by their nutrient status) such as gyres, upwelling, equatorial and coastal regions. Picocyanobacterial community structure analysis of the three previously unpublished cruise tracks is presented first, before we explore more general distribution patterns using the whole data set.

Eastern South Pacific Ocean (BIOSOPE cruise 2004)

Samples were collected during October–December 2004 on the BIOSOPE cruise which traversed an

Table 1. Details of cruises; for cruise tracks see Fig. 1.

Cruise	Date	Geographic region	Latitude/longitude	Sampling	Reference
1 – AMBITION	September 2001	Arabian Sea	1°S–26°N, 56–67°E	11 stations, 6 depths	Fuller *et al.* (2006)
2 – AMT15	September–October 2004	Atlantic Meridional Transect	48°N–40°S, 25°W–10°E	21 stations, 6 depths	Zwirglmaier *et al.* (2007)
3 – ARCTIC	August 2002	Arctic Ocean	70–79°N, 3–23°E	11 stations, 5 depths	This study
4 – BEAGLE	August 2003–January 2004	Circumnavigation of the southern hemisphere	20–32°S	15 stations, surface only	Bouman *et al.* (2006) and this study
5 – BIOSOPE	October–December 2004	Eastern South Pacific Ocean	8–34°S, 141–72°W	11 stations, 6 depths	This study
6 – VANC10MV	May–June 2003	Indian Ocean	35–12°S, 24–113°E	15 stations, 6 depths	This study
7 – PROSOPE	September–October 1999	Mediterranean Sea	31–43°N, 10°W–22°E	9 stations, 5 depths	Garczarek *et al.* (2007)
8 – RED SEA	February 1999–January 2000	Red Sea	29°N, 34°E	One station, 6 depths over an annual cycle	Fuller *et al.* (2005)

8000-km-long transect from subequatorial mesotrophic waters near to the Marquesas islands through hyperoligotrophic waters of the South Pacific gyre, considered to be some of the 'clearest' natural waters on Earth (Morel *et al.*, 2007), and extending to more eutrophic waters off the Chilean coastal upwelling near Concepción (Figs 1 and 2A).

Synechococcus abundance was low in the gyre (maximum 4×10^3 cells ml^{-1}), slightly higher in the subequatorial area (5×10^3 to 2×10^4 cells ml^{-1}) and highest in the upwelling region (up to 10^5 cells ml^{-1}), while *Prochlorococcus* cell numbers were highest adjacent to the hyperoligotrophic gyre area (up to 4×10^5 cells ml^{-1}), and slightly lower within the gyre (up to 1.6×10^5 cells ml^{-1}). Genetically, *Synechococcus* clades IV (up to 90% relative hybridization based on dot blot hybridization data) and I (up to 7% relative hybridization) dominated the upwelling region off Concepción, clade II was restricted to one station near the Marquesas while clades III and V/VI/VII were more broadly distributed along the transect (Fig. 2A). *Prochlorococcus* populations were dominated by the HLII lineage closest to the gyre centre giving a signal (up to 96% relative hybridization) that was two to three times higher and spatially separated from the HLI lineage whose maximum signal occurred adjacent to the hyperoligotrophic gyre area (up to 28% relative hybridization). LL *Prochlorococcus* lineages showed a particularly interesting distribution being fairly broadly distributed deep down in the water column throughout the cruise track, though with a greater signal at the eastern end of the transect (up to 60% relative hybridization signal) in transition waters between the South Pacific gyre and the coastal upwelling region.

Indian Ocean (VANC10MV cruise 2003)

Samples were collected on the VANC10MV cruise during May–June 2003 which travelled between Cape Town, South Africa and Port Hedland, Australia and passed through the south central Indian Ocean gyre (Figs 1 and 2b). *Synechococcus* cell numbers were up to 4×10^4 cells ml^{-1} in the coastal shelf area near Cape Town, dropping to 10^3 cells ml^{-1} in the gyre. In contrast, *Prochlorococcus* abundance peaked in the gyre with up to 2.5×10^5 cells ml^{-1}.

Synechococcus community structure comprised clade II genotypes in coastal waters near Cape Town (up to 16% relative hybridization in surface waters), whereas the clade V/VI/VII signal dominated further offshore (28% relative hybridization). Clades I and IV were barely detected anywhere along the cruise track (maximum 7% relative hybridization of clade IV, again in coastal waters off Cape Town, and generally with signals for both clades $< 1\%$ throughout the transect). For *Prochlorococcus*, HLI and HLII genotypes dominated the upper 50–100 m of the water column but were themselves strictly spatially separated, HLI being restricted to the western section of the transect and HLII the eastern section. LL genotypes were again confined to deeper waters, but the hybridization signal revealed a broader longitudinal distribution than either of the HL ecotypes and apparently horizontally separated 'between' the HLI and HLII signals (Fig. 2B).

Arctic Ocean

Samples were collected in an area between the Norwegian, Greenland and Barents Seas in two transects during

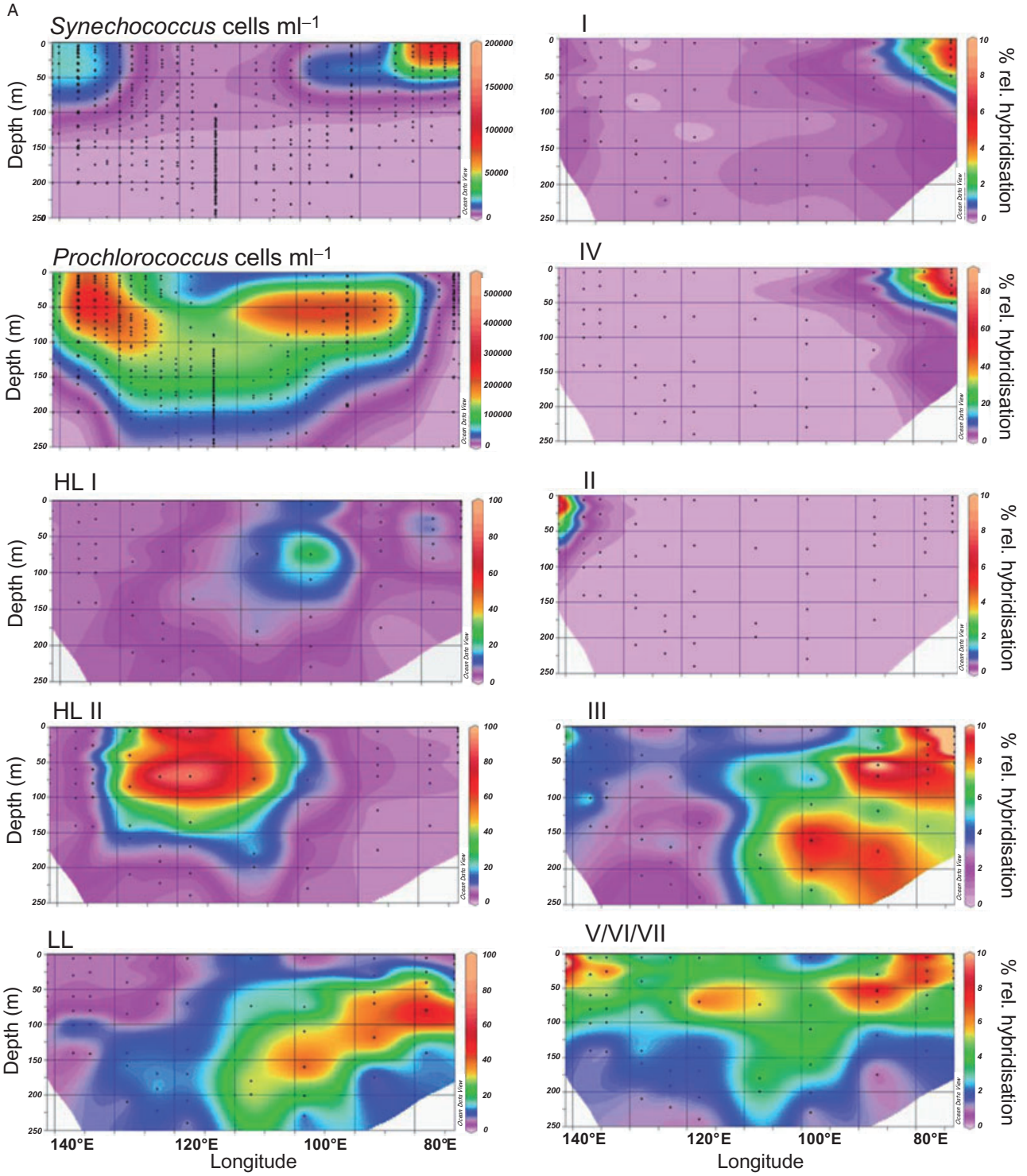

Fig. 2. Distribution of *Prochlorococcus* ecotypes and *Synechococcus* clades along the BIOSOPE (A), Indian Ocean (B) and Arctic Ocean (C) cruise tracks. Contour plots indicate percentage relative hybridization (as a proportion of all amplified by primers OXY107F and OXY1313R) of the different lineages, or abundance (cells ml^{-1}) based on flow cytometry data. Black dots represent the sampling points. For *Synechococcus*, only clades that reached relative hybridization values > 5% are shown. Note the differences in scale on the *y*-axis.

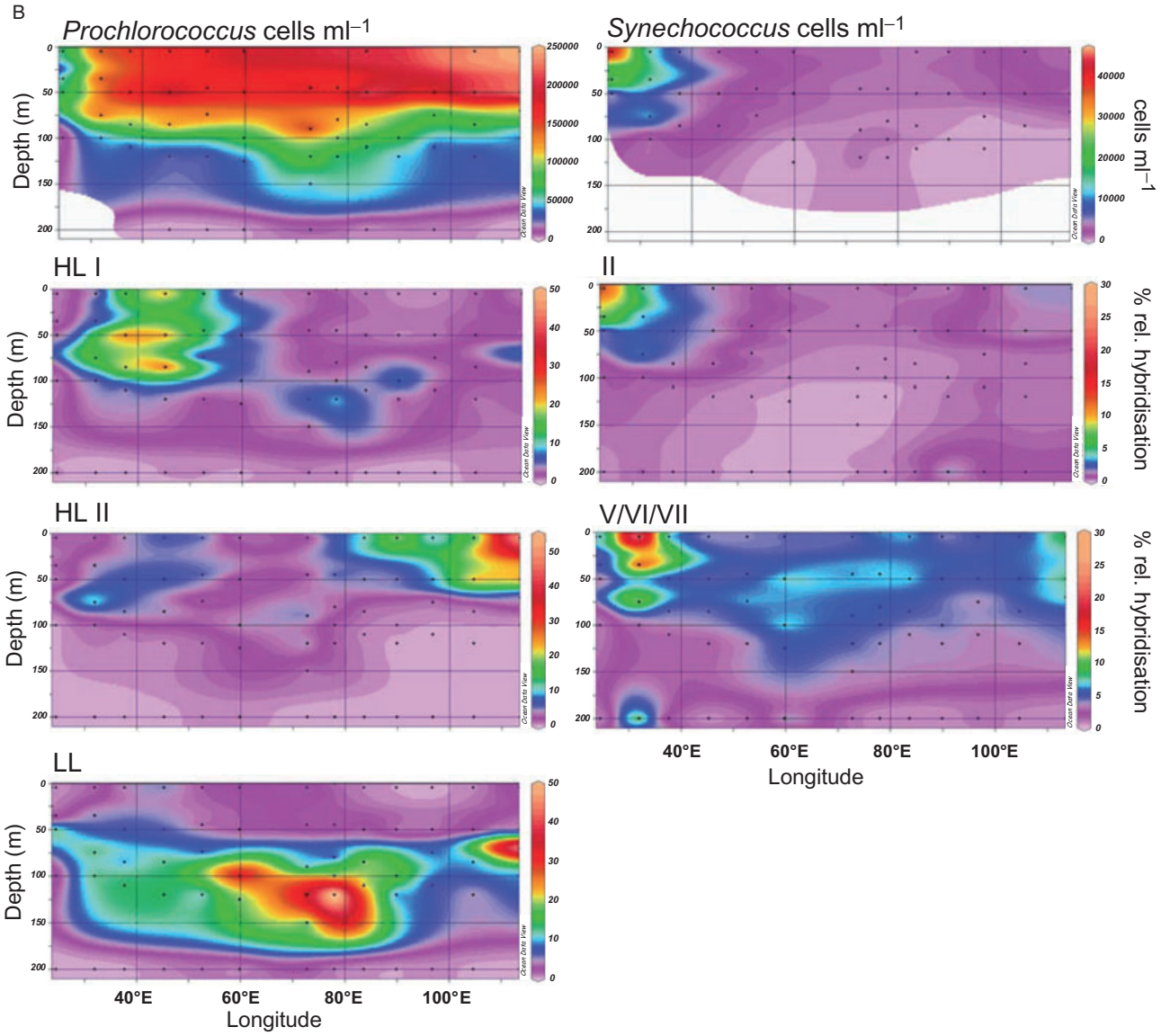

Fig. 2. *cont.*

August–September 2002 (Figs 1 and 2C). *Synechococcus* cell number ranged between 3×10^3 and 4×10^4 cells ml^{-1}. As expected, no *Prochlorococcus* was found in these high-latitude waters. The *Synechococcus* population was again dominated by clades I and IV, which co-occurred at the same stations (Fig. 2C). Interestingly, here clades I and IV gave approximately equal hybridization signals, whereas in the eastern South Pacific Ocean (see above) and Atlantic Ocean (Zwirglmaier *et al*., 2007), the clade IV signal was 5–10 times greater than clade I. Construction of clone libraries, and subsequent phylogenetic analysis at two stations along the transect that extended south to the Norwegian coastline (Fig. 2C, black triangles), confirmed the dominance of these lineages (14 of 15 sequences were either from clades I or IV) in this oceanic region.

Prochlorococcus *community structure: global insights*

Analysis of the combined data sets outlined in Fig. 1 revealed a general spatial partitioning of HL ecotypes, with HLI genotypes dominating more temperate latitudes (35°–48°N and 35°–40°S), and HLII restricted to subtropical and tropical regions (30°N–30°S), although there is some overlap at the transition between temperate and subtropical zones (Fig. 3). In contrast, LL *Prochlorococcus* ecotypes were relatively evenly distributed throughout 40°S–48°N although it should be noted that neither LLII

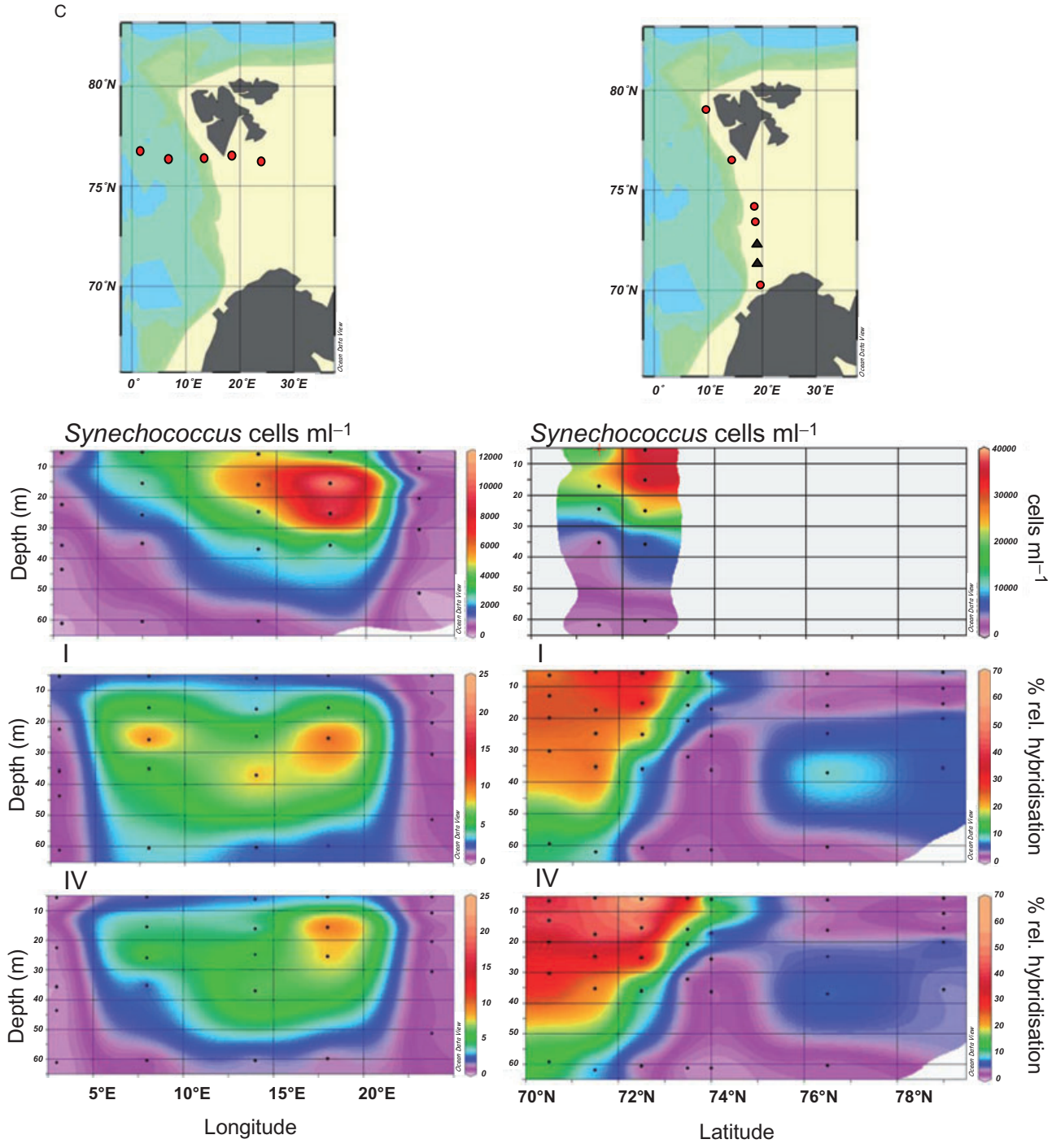

Fig. 2. *cont.*

(eSS120) or LLIII (eMIT9211) (see Ahlgren *et al.*, 2006b) is detected by our LL probe. A more detailed analysis with additional probes might reveal a distinct pattern for these strains.

Subsequent scatter plot analysis (Fig. 4A) reveals that temperature is a major determinant of this spatial partitioning of ecotypes, with the HLII ecotype distributed over a temperature range of 23–30°C, nicely reflecting their latitudinal distribution, while the ranges for the HLI and LL ecotypes are considerably lower (14–24°C and 14–25°C respectively). Laboratory culture work has previously also demonstrated this differential temperature optima for growth, certainly in HL ecotypes (Johnson *et al.*, 2006), while here we now show that earlier basin-scale studies

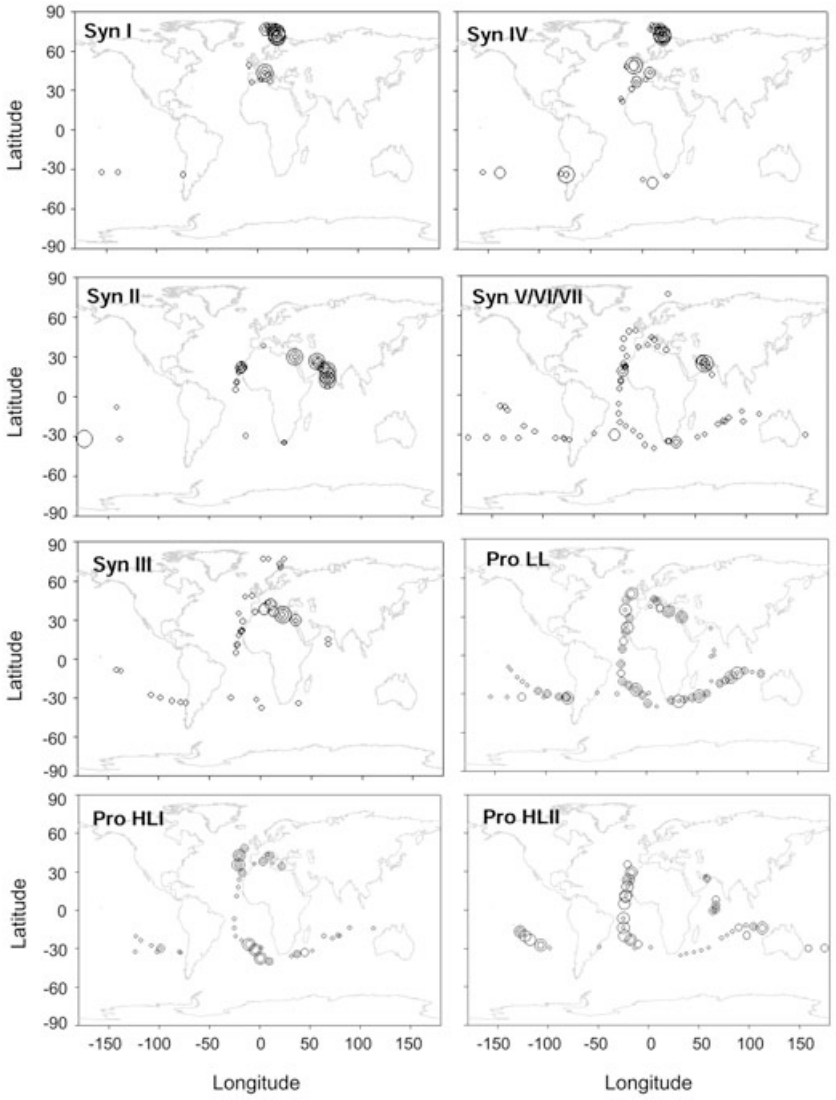

Fig. 3. Global distribution of *Synechococcus* clades and *Prochlorococcus* ecotypes analysed by dot blot hybridization. Size of the circle delineates the abundance of the lineage: small: 5–20%, medium: 20–50%, large: > 50% relative hybridization signal. Values < 5% are not shown.

where temperature has been implicated as a key ecological determinant dictating *Prochlorococcus* community structure (Bouman *et al*., 2006; Johnson *et al*., 2006; Zwirglmaier *et al*., 2007) can now be extended to the global scale.

Similarly, in addition to this horizontal spatial separation of HLI and HLII ecotypes, the 'global' data set (Fig. 4B) also reiterates the already well-documented observation of vertical separation of HL and LL ecotypes (see Partensky *et al*., 1999a; West and Scanlan, 1999; Johnson *et al*., 2006; Garczarek *et al*., 2007; Zwirglmaier *et al*., 2007) especially at relative hybridization values above 20%, and which equate to fully stratified water bodies. The reported exceptions to this paradigm, for example, of HLII genotypes extending to the base of the euphotic zone (West *et al*., 2001) or LL ecotypes dominating in surface waters likely equate to physically well-mixed water columns and are seen as outlying data points in Fig. 4B.

Synechococcus *community structure: global insights*

Specific *Synechococcus* lineages also show distinct distribution patterns at the global scale (Fig. 3). However, there is considerable variation in the abundance of specific lineages based on relative hybridization data. Clades I, II and IV were the only ones appearing in high abundance reaching levels up to 90% relative hybridization signal in different regions. Maximum signals for clade III (approximately 40% relative hybridization) and clades V/VI/VII (approximately 20% relative hybridization), the latter targeted by a single probe, suggest moderate abundance for these clades, while clades VIII, IX and X appear to be in low abundance as signals were always well below 5% relative hybridization.

Certainly the most striking observation was the co-occurrence and high relative hybridization signal of clades I and IV in the coastal boundary zone, reiterating

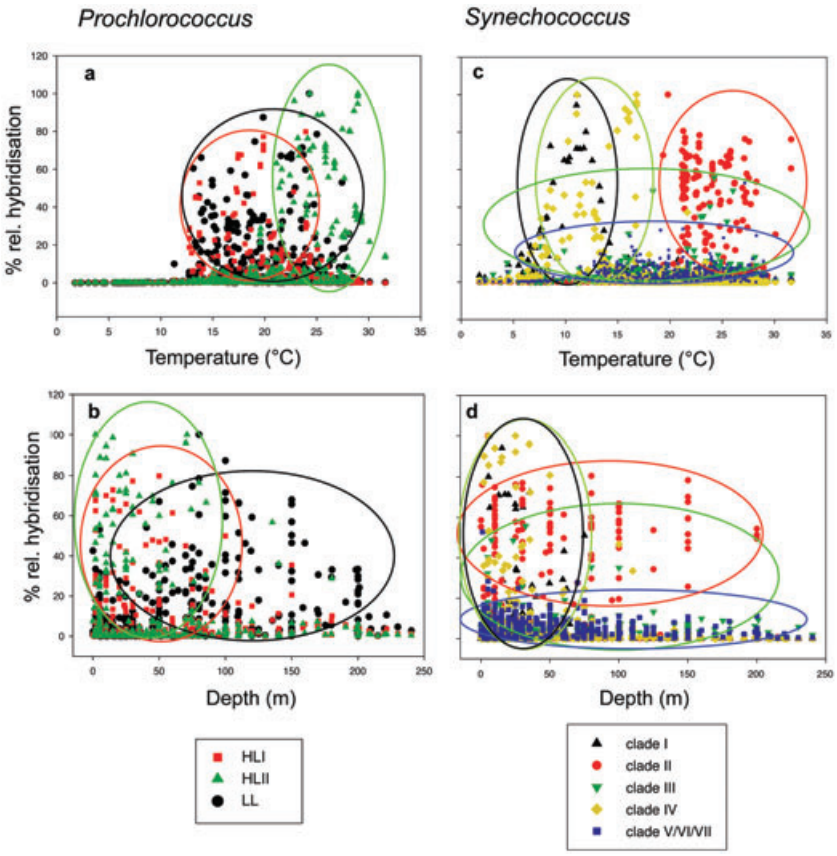

Fig. 4. Scatter plots of percentage relative hybridization values of *Prochlorococcus* ecotypes ($n = 439$ samples) and *Synechococcus* clades ($n = 526$) versus temperature and depth.

recent observations along an Atlantic Meridional Transect (AMT) (Zwirglmaier *et al*., 2007) and off the Californian coast (Brown *et al*., 2005). This co-occurrence was observed in eastern South Pacific Ocean, Mediterranean Sea and Arctic Ocean waters, but not in the Red Sea, Arabian Sea or Indian Ocean, where both clades I and IV were virtually absent. Thus, clades I and IV appear to be largely confined to higher latitudes, above *c*. 30°N and below 30°S, but not in the subtropical and tropical regions between. This distribution pattern may partly be explained by seawater temperature. Certainly, a scatter plot of clade I and IV relative hybridization values against temperature (Fig. 4C) shows the temperature range for clades I and IV to be very similar and clearly below that of other *Synechococcus* clades. Within their latitudinal range, clades I and IV are predominantly found in the coastal boundary zone, alongside a broad range of nitrate and phosphate concentrations (0.03–14.5 μM and 0.2–1.2 μM, respectively, data not shown).

Although there is a strong correlation of co-occurrence of clades I and IV (0.695, $P < 0.001$, $n = 525$), the ratio of I:IV varies considerably (Fig. 5). Interestingly, the ratio is constant within samples from a particular cruise, samples from AMT (Atlantic Ocean) and BIOSOPE (eastern South Pacific) have a low clade I:IV ratio, while the PROSOPE (Mediterranean Sea) and Arctic Ocean cruises have a higher ratio. A comparison of these cruises with regard to various physical and chemical parameters, e.g. latitude, temperature, nutrient concentration and time of the year sampled, did not point out any common factors for the cruises with the same I:IV clade ratio or differing factors for the cruises with differing ratios, suggesting that there are other 'unknown' factors determining the dominance of either clade I or IV in an environment. Recent genomic

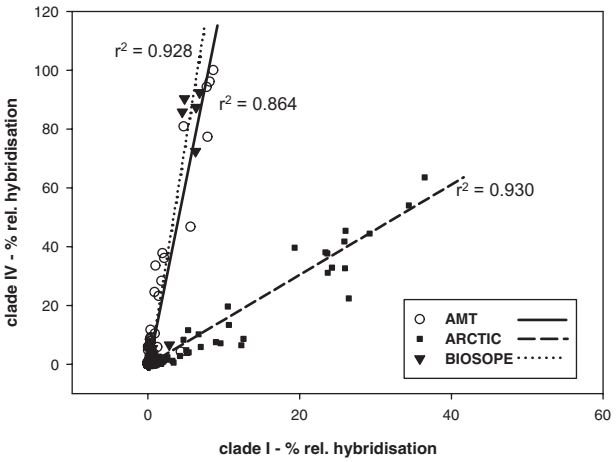

Fig. 5. Relative hybridization values of *Synechococcus* clade I plotted against clade IV for various cruises.

data for *Synechococcus* sp. CC9311, a clade I strain (Palenik *et al*., 2006), has revealed several features reflecting an adaptation to a coastal environment, such as a restricted phosphate regulatory and scavenging system as well as a greater capacity for uptake, use and storage of several metals, especially copper and iron. Interestingly, genomic data for *Synechococcus* sp. BL107 (a clade IV strain) show similar features (M. Ostrowski and D.J. Scanlan, unpubl. data). This raises the question, whether clades I and IV compete for the same ecological niche, or whether there are synergistic effects between the two promoting the growth of both together, but not without the other.

Synechococcus clade II may be the subtropical/tropical 'counterpart' of clades I and IV, as it appears in high abundance (up to 90% relative hybridization signal) usually in coastal/continental shelf zones, but in contrast to clades I and IV only in the subtropical/tropical latitudes between 30°S and 30°N. It is virtually absent above or below these latitudes (Fig. 3), although there is some overlap with clades I and IV in boundary waters (see also Ferris and Palenik, 1998). Similarly, the temperature profile for clade II (Fig. 4C), ranging between 22°C and 28°C, is considerably higher than that for either clade I or clade IV.

A comparison of the hybridization signal for the universal *Synechococcus* probe, SYN1258 (Fuller *et al*., 2003), with the sum of all lineage-specific probes in this global data set reveals that in the vast majority of samples $>95\%$ of the total *Synechococcus* signal is detected by the suite of specific probes for clades I–X (data not shown). However, there are some samples where a considerable fraction (10–30%) of *Synechococcus* remains unidentified. These samples are mostly coastal, in subtropical/tropical latitudes, and contain a high abundance of clade II genotypes based on relative hybridization data. Hence, it is possible that this as yet unidentified fraction contains a 'partner' clade to clade II, analogous to the clade I/IV relationship. Certainly, several new *Synechococcus* lineages have been described recently (Ahlgren and Rocap, 2006a; Penno *et al*., 2006), but as 16S rRNA gene sequence data for these clades are not yet available it is unclear whether they are detected by our current suite of 16S rRNA gene probes.

Clade II genotypes appear to have a broad depth distribution (Fig. 4D), from the surface down to *c.* 200 m, although most of these 'deep' measurements are derived from Red Sea seawater samples obtained during spring and winter, when the water column was mixed (Fuller *et al*., 2005). Hence, it is still possible that some clade II *Synechococcus* strains are more abundant in surface layers of the euphotic zone as has been reported previously (Toledo and Palenik, 2003). Certainly there was no obvious depth partitioning for any of the major *Synechococcus* lineages in the global data set, although the major abundance of clades I and IV was often in surface waters (Fig. 4D).

In contrast to clades I, II and IV, clades III and V/VI/VII showed no obvious latitudinal preference (Fig. 3). However, clade III genotypes were confined to a fairly narrow window of nitrate and phosphate macronutrient concentration (data not shown) suggesting members of this clade are oligotrophs. Certainly, genomic data for the clade III isolate *Synechococcus* sp. WH8102 (Palenik *et al*., 2003), which show it contains a plethora of nutrient transporters but little regulatory machinery, also point to occupation of a stable, low-nutrient environment. Clades V, VI and VII genotypes appear to be widely distributed in oceanic waters, perhaps suggesting members of these clades are 'generalists' or 'opportunists'. However, as these clades are detected only by a single 16S rRNA gene oligonucleotide probe, some caution needs to be placed here. Use of a less conserved marker gene, e.g. *rpoC1* or ITS, in future molecular ecological studies will help to discern whether the individual lineages transgress from this 'general' distribution and confer specific distribution patterns.

Picocyanobacterial community structure and oceanic biomes

Four major ocean domains or biomes have been reported for oceanic systems: the polar domain, coastal boundary domain, westerly winds domain and trade winds domain (Longhurst, 1995; 2007; see Fig. 1). These domains differ in their temperature range, nutrient concentration and degree of seasonal changes in mixing, light conditions and primary production. Comparing the abundance of individual *Synechococcus* and *Prochlorococcus* lineages reveals a distinct pattern or 'fingerprint' distribution for each domain (Fig. 6). The polar domain (Fig. 6A) is dominated by *Synechococcus* clades I and IV, with other *Synechococcus* lineages appearing only in low abundance while *Prochlorococcus*, due to the high latitudes of this biome, are completely absent. The coastal boundary domain (Fig. 6B), on the other hand, is dominated by *Synechococcus* clade II, although sporadically high abundance of other lineages, especially *Synechococcus* clade IV, can appear. The trade winds domain (subtropical and tropical latitudes) (Fig. 6C) and westerly winds domain (temperate latitudes) (Fig. 6D) both show a similar and rather balanced distribution of all *Synechococcus* lineages in low to moderate abundance. They are clearly dominated by *Prochlorococcus*, but whereas HLII ecotypes are more prevalent in the trade winds domain, HLI ecotypes predominate in the westerly winds domain. The weak seasonality of the trade winds domain and

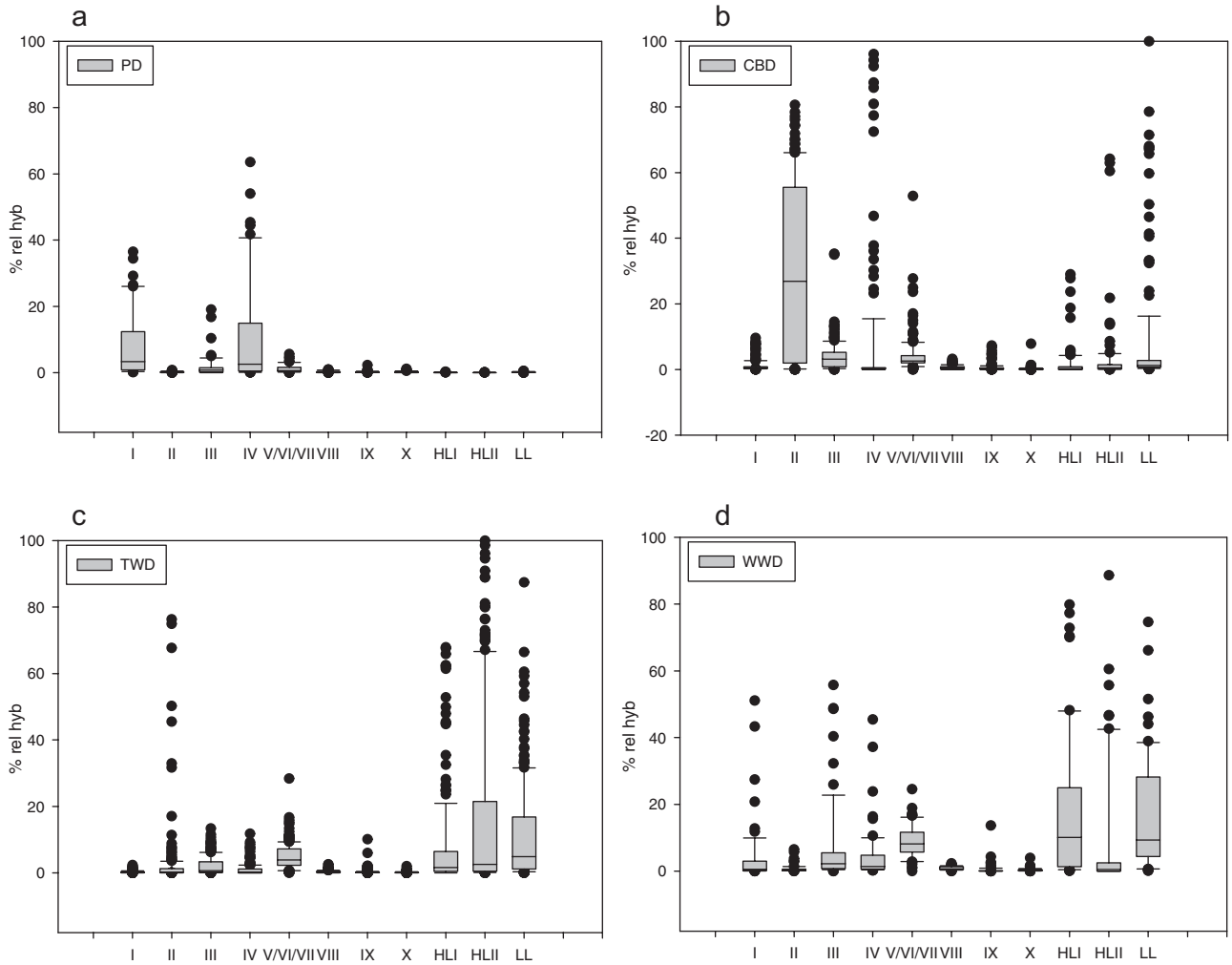

Fig. 6. Box and whisker plots of *Synechococcus* and *Prochlorococcus* lineage distribution in the four major ocean domains. (A) Polar domain (PD), (B) coastal boundary domain (CBD), (C) trade winds domain (TWD), (D) westerly winds domain (WWD). The box represents the interquartile range including the median, the ends of the whiskers represent the 10th and 90th percentile, outliers > 90% and < 10% are shown as dots. Number of samples analysed in each domain: PD: $n = 55$, WWD: $n = 63$, TWD: $n = 228$, CBD: $n = 179$.

hence relatively constant mixed-layer depth of this biome, in contrast to the strongly seasonal mixed-layer depth of the westerlies biome (Longhurst, 2007), likely underlies this observation. Certainly, it is entirely consistent with HLII genotypes dominating surface waters of regions with high stratification, whereas HLI is more prevalent in surface waters with moderate stratification and mixed layer depth (Bouman *et al*., 2006).

Multivariate analysis – correlations of lineage abundance to environmental parameters

Factor analysis was used to analyse the relative hybridization patterns of the different *Synechococcus* and *Prochlorococcus* lineages (Fig. 7). Separate analyses were performed using data from different geographic areas: the Atlantic Ocean (AMT15 cruise, Fig. 7A), eastern South Pacific (BIOSOPE cruise, Fig. 7B) and a 'global' data set using data from the AMT, AMBITION (Arabian Sea), BIOSOPE, PROSOPE (Mediterranean Sea), BEAGLE and Arctic Ocean (Fig. 7C). The results for these three data sets are remarkably similar in that (i) each discriminates *Prochlorococcus* and *Synechococcus* lineages into two clearly separable groups on axis 1, presumably reflecting their observed largely complementary distribution patterns in the environment, (ii) *Prochlorococcus* ecotypes form a rather tight cluster, while *Synechococcus* clades show a broader discrimination, either on axis 1 or axis 2, perhaps a function of a broader physiological diversity in the latter genus, and (iii) *Synechococcus* clades I, IV on axis 1 and 2 and *Synechococcus* clade II on axis 1, which occur in high abundance in specific areas, occupy extreme positions in the graph.

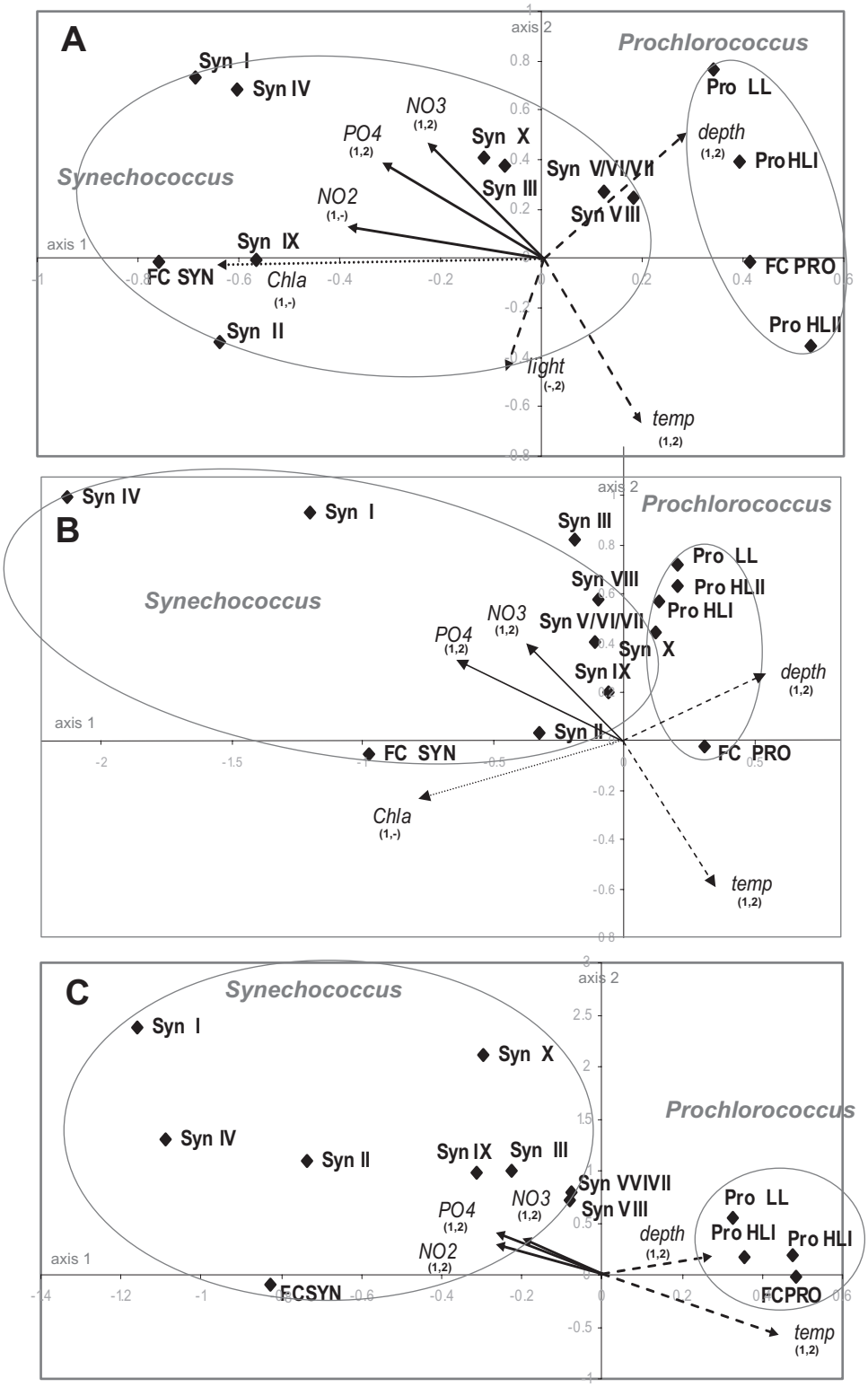

Fig. 7. Factor analysis of (A) AMT ($n = 110$ samples), (B) BIOSOPE ($n = 64$ samples) and (C) global (including AMT, BIOSOPE, PROSOPE, AMBITION, BEAGLE and Arctic Ocean, $n = 314$ samples) data sets. Arrows represent correlation coefficients between environmental parameters and the first two ordination axes. Circles define groups of *Synechococcus* clades and *Prochlorococcus* ecotypes. Solid arrows represent chemical parameters and dashed arrows represent physical parameters. Numbers adjacent to these parameters indicate a significant correlation with axis 1 and/or axis 2 ($P < 0.05$); a dash indicates no significant correlation.

Table 2. Variation partition analysis.

	Prochlorococcus	*Synechococcus*
% of variance explained by physical parameters	40.7	7.4
% of variance explained by chemical parameters	6.8	22.4
Shared	10.6	6.8
Unexplained	41.9	63.4

A correlation analysis between the two first ordination axes and explanatory variables was performed to understand the main factors controlling the distribution of the two picocyanobacterial genera (arrows in Fig. 7). Arrows represent correlation coefficients between the explanatory variables and the first two ordination axes. *Prochlorococcus* appears to be more influenced by the physical parameters temperature and depth (Fig. 7C, dashed arrows) and *Synechococcus* by nutrients (PO_4^{3-}, NO_2^- and NO_3^-; Fig. 7C, solid arrows). This last observation was further confirmed by variation partition analysis (Muylaert *et al*., 2002; Jardillier *et al*., 2005), which showed that for *Prochlorococcus* 40.7% of the variance can be significantly explained by the physical parameters temperature, light and depth, while nutrients (PO_4^{3-}, NO_2^- and NO_3^-) account for 22.4% of the variance of *Synechococcus* lineages (Table 2). It is striking that, for both genera, *Prochlorococcus* but especially *Synechococcus*, a large part of the variation remains unexplained (48% and 64% respectively) by the parameters analysed here, even though these are the parameters that are most commonly considered when attempting to explain spatial partitioning patterns. Clearly, other physical and biological factors must be taken into account to understand the observed community structure of these organisms.

Certainly missing from our data set are any trace metal measurements, and as marine picocyanobacteria show differential toxicity to copper (Mann *et al*., 2002), or conversely potential limitation for other metals, e.g. cobalt/iron (Mann and Chisholm, 2000; Saito *et al*., 2002; 2005), then further information on *in situ* trace metal chemistry and associated trace metal physiology of these organisms is clearly needed. Moreover, there are relatively few data focused towards understanding how biotic factors, e.g. grazing, phage infection or competition, affect community structure. Indeed, simple differences in cell size between the two genera and concomitant effects on nutrient uptake and/or grazing potential may likely explain some of this variation. Furthermore, biotic and abiotic factors may not be mutually exclusive given that copper has been shown to cause induction of a temperate marine cyanophage (Sode *et al*., 1997) while phosphate limitation has been implicated in inducing pseudolysogeny in marine *Synechococcus*, in which a phage-infected cell grows and divides even though its virus is pursuing a lytic infection (Mann, 2003).

Conclusions

Spatial partitioning of specific picocyanobacterial lineages at the global scale resulting in distinct signatures for specific ocean biomes is a basic requirement for predictable modelling of bacterial community structure in marine ecosystems. That this large-scale partitioning is evident even though there is likely microscale heterogeneity in environmental niche space that can give rise to extensive genotypic diversity within bacterial populations (see Thompson *et al*., 2005; Rusch *et al*., 2007) suggests there is still a lot to learn, however, about the general principles, particularly at the genomic level, governing bacterial population structure. Whether this involves a dynamic interplay between ecology and evolution within communities, as has recently been suggested (Johnson and Stinchcombe, 2007), awaits further work.

Experimental procedures

Sampling

Environmental samples ($n = 525$) were collected on several cruises between 1998 and 2004 (summarized in Table 1) and processed as described previously (Zwirglmaier *et al*., 2007). Flow cytometric analyses for the BIOSOPE, Indian Ocean and Arctic Ocean cruises were performed as described in Grob and colleagues (2007), F. Not, M. Latasa, R. Scharek, M. Viprey, P. Karleskind, I. Ontoria, A. Cumino, E. Goetze, D. Vaulot, and R. Massana (submitted) and Not and colleagues (2005) respectively.

Dot blot hybridization

Dot blot hybridization was carried out as described previously (Fuller *et al*., 2003; Zwirglmaier *et al*., 2007) using a set of 16S rRNA-targeted probes specific for *Synechococcus* clades I–X (Fuller *et al*., 2003) and *Prochlorococcus* ecotypes HLI, HLII and LL (West and Scanlan, 1999; Fuller *et al*., 2005). Relative hybridization for each sample represents the signal of the picocyanobacterial lineage-specific probe as a proportion of the total oxygenic phototroph 16S rRNA gene sequences amplified by the OXY107F–OXY1313R primer pair.

Clone library construction and DNA sequencing

16S rRNA gene clone libraries were constructed from two stations along the Arctic Ocean cruise track which showed

high abundance of *Synechococcus* at 5 m depth (Fig. 2C, black triangles). The 16S rRNA gene was amplified using the polymerase chain reaction (PCR) with oxygenic phototroph specific primers OXY107F and OXY1313R (West *et al*., 2001) and cloned into *Escherichia coli* using the Invitrogen TOPO TA cloning kit (Invitrogen, USA). Polymerase chain reaction products were sequenced bidirectionally using Big Dye Terminator Version 3.1 Chemistry (Applied Biosystems, Foster City, CA) and run on the 3100 Genetic Analyser. The sequences reported in this article have been deposited in the GenBank database under the following Accession No. EF622489 to EF622503.

Statistical analyses

A matrix of *Synechococcus* and *Prochlorococcus* flow cytometry cell counts, and percentage relative hybridization data for individual picocyanobacterial lineages was subjected to factor analysis using the ADE4 software package (http://cran.r-project.org/). Correlation analysis between the two ordination axes and explanatory (environmental) variables was performed to understand the main factors controlling picocyanobacterial lineage distribution patterns. Explanatory variables were $\log(n+1)$ transformed prior to analysis; response variables (clade abundance) were square root transformed.

To evaluate the effects of physical and chemical parameters on *Prochlorococcus* and *Synechococcus* abundance, multivariate analysis with variation partitioning (variation partitioning analysis or VPA) was used (see Borcard *et al*., 1992; Muylaert *et al*., 2002). All explanatory variables were divided into two groups: chemical variables (PO_4^{3-}, NO_2^- and NO_3^-) and physical variables (temperature, depth, light). For each experiment, we selected only variables that independently explained a significant proportion ($P < 0.05$) of the variation in *Prochlorococcus* and *Synechococcus* abundance by forward canonical correspondence analysis (CCA). For both CCA and VPA analyses, the vegan package within the R software was used (http://cran.r-project.org/). Then, for the set of chemical and physical variables separately, we generated a minimal set of explanatory variables explaining variation in cyanobacterial abundance. Variation partitioning analysis allowed us to distinguish purely chemical or physical effects on *Prochlorococcus* and *Synechococcus* community structure as well as the part explained by both these effects (i.e. the shared part).

Acknowledgements

The authors thank the chief scientists, captains and crews of all the research vessels involved in obtaining data used in this study. Cruise work was supported by (i) the French PROOF programmes PROSOPE and BIOSOPE as well as the regional programme IMPALA (Région Bretagne), (ii) Spanish grants REN2001-4909-E/ANT and REN2002-10951-E/MAR and (iii) the Chilean National Commission for Scientific and Technological Research through the FONDAP programme, and by the Millennium Scientific Initiative through Grant EMBA P04/007. This study was supported by the UK Natural Environment Research Council via Grant NER/A/S/2002/00795 to D.J.S. Additional funding was also obtained via the ANR PICOFUNPAC project.

References

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

Ahlgren, N.A., Rocap, G., and Chisholm, S.W. (2006b) Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environ Microbiol* **8:** 441–454.

Borcard, D., Legendre, P., and Drapeau, P. (1992) Partialling out the spatial component of ecological variation. *Ecology* **73:** 1045–1055.

Bouman, H.A., Ulloa, O., Scanlan, D.J., Zwirglmaier, K., Li, W.K.W., Platt, T., *et al.* (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312:** 918–921.

Brown, M.V., Schwalbach, M.S., Hewson, I., and Fuhrman, J.A. (2005) Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ Microbiol* **7:** 1466–1479.

Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., and Welschmeyer, N.A. (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334:** 340–343.

Chisholm, S.W., Frankel, S.L., Goericke, R., Olson, R.J., Palenik, B., Waterbury, J.B., *et al.* (1992) Prochlorococcus marinus nov. gen. nov. sp. an oxyphototrophic marine prokaryote containing divinyl chlorophyll a and b. *Arch Microbiol* **157:** 297–300.

Dolan, J.R. (2005) Biogeography of aquatic microbes. *Aquat Microb Ecol* **41:** 39–48.

Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I.M., Barbe, V., *et al.* (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci USA* **100:** 10020–10025.

Ferris, M.J., and Palenik, B. (1998) Niche adaptation in ocean cyanobacteria. *Nature* **396:** 226–228.

Follows, M.J., Dutkiewicz, S., Grant, S., and Chisholm, S.W. (2007) Emergent biogeography of microbial communities in a model ocean. *Science* **315:** 1843–1846.

Fuller, N.J., Marie, D., Partensky, F., Vaulot, D., Post, A.F., and Scanlan, D.J. (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.

Fuller, N.J., West, N.J., Marie, D., Yallop, M., Rivlin, T., Post, A.F., and Scanlan, D.J. (2005) Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol Oceanogr* **50:** 363–375.

Fuller, N.J., Tarran, G.A., Yallop, M., Orcutt, K.M., and Scanlan, D.J. (2006) Molecular analysis of picocyanobacterial community structure along an Arabian Sea transect reveals distinct spatial separation of lineages. *Limnol Oceanogr* **51:** 2515–2526.

Garczarek, L., Dufresne, A., Rousvoal, S., West, N.J., Mazard, S., Marie, D., *et al.* (2007) High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol* **60:** 189–206.

Grob, C., Ulloa, O., Claustre, H., Huot, Y., Alarcón, G., and Marie, D. (2007) Contribution of picoplankton to the total particulate organic carbon (POC) concentration in the eastern South Pacific. *Biogeosci Discuss* **4:** 1–37.

Herdman, M., Castenholz, R.W., Iteman, I., Waterbury, J., and Rippka, R. (2001) The Archaea and the deeply branching and phototrophic bacteria. Subsection I (formerly Chroococcales, Wettstein 1924, emend. Rippka, Deruelles, Waterbury, Herdman and Stanier 1979). In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, Vol. 1. Boone, D.R., and Castenholz, R.W. (eds). New York, USA: Springer-Verlag, pp. 493–514.

Heywood, J.L., Zubkov, M.V., Tarran, G.A., Fuchs, B.M., and Holligan, P.M. (2006) Prokaryoplankton standing stocks in oligotrophic gyre and equatorial provinces of the Atlantic Ocean: evaluation of inter-annual variability. *Deep Sea Res Part II: Top Stud Oceanogr* **53:** 1530–1547.

Jardillier, L., Boucher, D., Personnic, S., Jacquet, S., Thénot, A., Sargos, D., *et al.* (2005) Relative importance of nutrients and mortality factors on prokaryotic community composition in two lakes of different trophic status: microcosm experiments. *FEMS Microbiol Ecol* **53:** 429–443.

Johnson, M.T.J., and Stinchcombe, J.R. (2007) An emerging synthesis between community ecology and evolutionary biology. *Trends Ecol Evol* **22:** 250–257.

Johnson, Z.I., Zinser, E.R., Coe, A., McNulty, N.P., Woodward, E.M.S., and Chisholm, S.W. (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311:** 1737–1740.

Li, W.K.W. (1994) Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39:** 169–175.

Liu, H., Nolla, H.A., and Campbell, L. (1997) *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquat Microb Ecol* **12:** 39–47.

Longhurst, A. (1995) Seasonal cycles of pelagic production and consumption. *Prog Oceanogr* **36:** 77–167.

Longhurst, A. (2007) *Ecological Geography of the Sea*, 2nd edn. London, UK: Academic Press.

Mann, E.L., and Chisholm, S.W. (2000) Iron limits the cell division rate of *Prochlorococcus* in the eastern equatorial Pacific. *Limnol Oceanogr* **45:** 1067–1076.

Mann, E.L., Ahlgren, N., Moffett, J.W., and Chisholm, S.W. (2002) Copper toxicity and cyanobacteria ecology in the Sargasso Sea. *Limnol Oceanogr* **47:** 976–988.

Mann, N.H. (2003) Phages of the marine cyanobacterial picophytoplankton. *FEMS Microbiol Rev* **27:** 17–34.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

Moore, L.R., and Chisholm, S.W. (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates. *Limnol Oceanogr* **44:** 628–638.

Morel, A., Gentili, B., Claustre, H., Babin, M., Bricaud, A., Ras, J., and Tièche, F. (2007) Optical properties of the 'clearest' natural waters. *Limnol Oceanogr* **52:** 217–229.

Muhling, M., Fuller, N.J., Somerfield, P.J., Post, A.F., Wilson, W.H., Scanlan, D.J., *et al.* (2006) High resolution genetic diversity studies of marine *Synechococcus* isolates using *rpoC1*-based restriction fragment length polymorphism. *Aquat Microb Ecol* **45:** 263–275.

Muylaert, K., van der Gucht, K., Vloemans, N., de Meester, L., Gillis, M., and Vyverman, W. (2002) Relationship between bacterial community composition and bottom-up versus top-down variables in four eutrophic shallow lakes. *Appl Environ Microbiol* **68:** 4740–4750.

Not, F., Massana, R., Latasa, M., Marie, D., Colson, C., Eikrem, W., *et al.* (2005) Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* **50:** 1677–1686.

Palenik, B., Brahamsha, B., Larimer, F.W., Land, M., Hauser, L., Chain, P., *et al.* (2003) The genome of a motile marine *Synechococcus*. *Nature* **424:** 1037–1042.

Palenik, B., Ren, Q., Dupont, C.L., Myers, G.S., Heidelberg, J.F., Badger, J.H., *et al.* (2006) Genome sequence of *Synechococcus* CC9311: insights into adaptation to a coastal environment. *Proc Natl Acad Sci USA* **103:** 13555–13559.

Partensky, F., Hess, W.R., and Vaulot, D. (1999a) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63:** 106–127.

Partensky, F., Blanchot, J., and Vaulot, D. (1999b) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. In *Marine Cyanobacteria*. No. spécial 19. Charpy, L., and Larkum, A.W.D. (eds). Monaco: Bulletin de l'Institut oceanographique, pp. 457–475.

Penno, S., Lindell, D., and Post, A.F. (2006) Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ Microbiol* **8:** 1200–1211.

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

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424:** 1042–1047.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al.* (2007) The *Sorcerer II* global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77.

Saito, M.A., Moffett, J.W., Chisholm, S.W., and Waterbury, J.B. (2002) Cobalt limitation and uptake in *Prochlorococcus*. *Limnol Oceanogr* **47:** 1629–1636.

Saito, M.A., Rocap, G., and Moffett, J.W. (2005) Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol Oceanogr* **50:** 279–290.

Scanlan, D.J., and West, N.J. (2002) Molecular ecology of the

marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiol Ecol* **40:** 1–12.

Sode, K., Oonari, R., and Oozeki, M. (1997) Induction of a temperate marine cyanophage by heavy metal. *J Mar Biotech* **5:** 178–180.

Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J., *et al.* (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307:** 1311–1313.

Ting, C.S., Rocap, G., King, J., and Chisholm, S.W. (2002) Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol* **10:** 134–142.

Toledo, G., and Palenik, B. (2003) A *Synechococcus* serotype is found preferentially in surface marine waters. *Limnol Oceanogr* **48:** 1744–1755.

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.

Veldhuis, M.J.W., Kraay, G.W., van Bleijswijk, J.D.L., and Baars, M.A. (1997) Seasonal and spatial variability in phytoplankton biomass, productivity and growth in the northwestern Indian ocean: the southwest and northeast monsoon, 1992–1993. *Deep-Sea Res Part I: Oceanogr Res Pap* **44:** 425–449.

West, N.J., and Scanlan, D.J. (1999) Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Appl Environ Microbiol* **65:** 2585–2591.

West, N.J., Schönhuber, W.A., Fuller, N.J., Amann, R.I., Rippka, R., Post, A.F., and Scanlan, D.J. (2001) Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by *in situ* hybridization using 16S rRNA-targeted oligonucleotides. *Microbiology* **147:** 1731–1744.

Zwirglmaier, K., Heywood, J.L., Chamberlain, K., Woodward, E.M.S., Zubkov, M.V., and Scanlan, D.J. (2007) Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. *Environ Microbiol* **9:** 1278–1290.

