

High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer

Laurence Garczarek¹, Alexis Dufresne¹, Sylvie Rousvoal¹, Nyree J. West², Sophie Mazard², Dominique Marie¹, Hervé Claustre³, Patrick Raimbault⁴, Anton F. Post⁵, David J. Scanlan² & Frédéric Partensky¹

¹Station Biologique, UMR 7144 CNRS & Université Pierre et Marie Curie, Roscoff, France; ²Department of Biological Sciences, University of Warwick, Coventry, UK; ³Laboratoire d'Océanographie de Villefranche, UMR 7093 CNRS & Université Pierre et Marie Curie, Villefranche-sur-mer, France; ⁴Laboratoire d'Océanographie et de Biogéochimie, UMR 6535 CNRS & Université Aix-Marseille 2, Marseille, France; and ⁵Interuniversity Institute of Marine Science, The H. Steinitz Marine Biology Lab, Eilat, Israel

Correspondence: Laurence Garczarek, Station Biologique, UMR7144 CNRS & Université Pierre et Marie Curie, Roscoff, France. Tel.: +33 2 98 29 25 64; fax: +33 2 98 29 23 24; e-mail: garczarek@sb-roscoff.fr

Present address: Nyree J. West, Observatoire Océanologique, UMR 7621 CNRS & Université Pierre et Marie Curie, BP 44, F-66651 Banyuls-sur-mer, France.

Received 30 June 2006; revised 4 October 2006; accepted 12 December 2006.
First published online April 2007.

DOI:10.1111/j.1574-6941.2007.00297.x

Editor: Patricia Sobecky

Keywords

Prochlorococcus; Mediterranean sea; ecotype; light harvesting complex; IsiA.

Introduction

The prokaryotic component of marine phytoplankton is dominated by two main genera: *Prochlorococcus* and *Synechococcus*. *Synechococcus* is the most ubiquitous, as cells of this genus can be found in virtually all environments from the equator up to latitudes of 60° and more (Liu *et al.*, 2002), whereas *Prochlorococcus* is restricted to a narrower latitudinal band, including most temperate and tropical waters (Buck *et al.*, 1996; Partensky *et al.*, 1999a, b; Johnson *et al.*, 2006). However, on a global scale, *Prochlorococcus* is approximately three times more abundant than *Synechococcus* (Partensky *et al.*, 1999a, b; Garcia-Pichel *et al.*, 2003). Indeed, *Prochlorococcus* cell densities vary only over a limited range along horizontal nutrient gradients, whereas *Synechococcus* concentrations can vary over more than two orders of magnitude along such gradients (Olson *et al.*, 1990; Partensky *et al.*, 1996). Furthermore, in warm oligo-

Abstract

Natural populations of the marine cyanobacterium *Prochlorococcus* exist as two main ecotypes, inhabiting different layers of the ocean's photic zone. These so-called high light- (HL-) and low light (LL-) adapted ecotypes are both physiologically and genetically distinct. HL strains can be separated into two major clades (HLI and HLII), whereas LL strains are more diverse. Here, we used several molecular techniques to study the genetic diversity of natural *Prochlorococcus* populations during the Prosope cruise in the Mediterranean Sea in the summer of 1999. Using a dot blot hybridization technique, we found that HLI was the dominant HL group and was confined to the upper mixed layer. In contrast, LL ecotypes were only found below the thermocline. Secondly, a restriction fragment length polymorphism analysis of PCR-amplified *pcb* genes (encoding the major light-harvesting proteins of *Prochlorococcus*) suggested that there were at least four genetically different ecotypes, occupying distinct but overlapping light niches in the photic zone. At comparable depths, similar banding patterns were observed throughout the sampled area, suggesting a horizontal homogenization of ecotypes. Nevertheless, environmental *pcb* gene sequences retrieved from different depths at two stations proved all different at the nucleotide level, suggesting a large genetic microdiversity within those ecotypes.

trophic waters, *Prochlorococcus* populations extend much deeper than *Synechococcus* (Campbell *et al.*, 1994; Partensky *et al.*, 1999a, b). Observation of bimodal populations in red fluorescence plots suggested that two distinct *Prochlorococcus* subpopulations coexist in the water column, occupying different but overlapping light niches (Campbell & Vaulot, 1993). This was confirmed by Moore *et al.* (1998), who isolated both fluorescence types by cell sorting and were able to grow them separately. Strains representative of populations from the upper and lower parts of the photic layer differed both physiologically, by their different growth irradiance optima, maximum photosynthetic yield and divinyl-chlorophyll (DV-Chl) *a* to *b* ratios, and genetically, as they clustered apart in 16S rRNA gene trees. Therefore it was proposed that these populations corresponded to different 'ecotypes', i.e. closely related genotypes that are adapted to specific ecological niches and that have a characteristic ecophysiology (Moore *et al.*, 1998; Moore &

Chisholm, 1999). The extent of genetic differences among strains of the so-called low light (hereafter LL)-adapted clades is much larger than between the two main high light (hereafter HL)-adapted clades identified to date, often called HLI and HLII (West & Scanlan, 1999; West *et al.*, 2001; Rocap *et al.*, 2002, 2003). This is likely due to the fact that HL clades are evolutionary more recent than LL clades (Dufresne *et al.*, 2005).

The photosynthetic apparatus of *Prochlorococcus* spp., when compared to that of marine *Synechococcus*, shows a number of genus-specific traits including the replacement of the light-harvesting phycobilisome complexes – their biosynthesis and regulation involves at least 29 different genes in *Synechococcus* sp. WH8102 (Six *et al.*, 2005) – by membrane-intrinsic divinyl-Chl *a/b*-binding proteins, encoded by *pcb* genes (La Roche *et al.*, 1996; Partensky & Garczarek, 2003). The number of *pcb* genes is variable (from one to eight) between *Prochlorococcus* strains (Garczarek *et al.*, 2000, 2001; Bibby *et al.*, 2003). The sole *pcb* gene of the HLI strain MED4 encodes a protein which serves as an antenna for photosystem II (PS II) only, whereas the LL-adapted strain SS120, which possesses eight different *pcb* genes, has antenna proteins for both PS I and for PS II (Bibby *et al.*, 2003). This multiplication of antenna proteins in some LL strains may correspond to an optimization strategy for photon capture, consistent with the reduced photon flux at depths. However, the LL-adapted strain MIT9313, located at the base of the *Prochlorococcus* radiation in 16S rRNA gene trees (Moore *et al.*, 1998; Rocap *et al.*, 2003), has only two *pcb* genes, one coding for a constitutive PSII antenna and the other for a PSI antenna induced only under iron deficiency (Bibby *et al.*, 2003). The latter is analogous to the IsiA protein-pigment complexes found in typical cyanobacteria (Bibby *et al.*, 2001; Boekema *et al.*, 2001). Surprisingly, a similar set of two distinct *pcb* genes is also found in the HLII strain MIT9312 (http://genome.jgi-psf.org/finished_microbes/pro_2/pro_2.home.html).

Differences in *pcb* copy number between LL and HL strains but also variability among known *pcb* sequences, even among the multiple genes of *P. marinus* SS120, allowed us to use a simple restriction fragment length polymorphism (RFLP) method to differentiate the genetic signature of cultured strains (Garczarek *et al.*, 2000, 2001). In the present study, we have applied this method, together with dot blot hybridization using 16S rRNA gene probes and *pcb* gene clone library sequencing to analyze the diversity of natural *Prochlorococcus* populations of the Mediterranean Sea in summer. This complementary suite of molecular analyses gave us access to different levels of genetic diversity within the *Prochlorococcus* genus in the environmental conditions studied here (i.e. strongly stratified waters). The first two techniques allowed us to look at the diversity of 'ecotypes'; RFLP analysis of *pcb* gene amplicons allowed a much better

resolution than dot blot hybridization, but in return was more difficult to relate to known ecotypes. In contrast, sequencing of *pcb* gene clone libraries permitted us to look at the *Prochlorococcus* 'genotype' diversity level. This combination of approaches brought unprecedented insights into the diversity and distribution of the *Prochlorococcus* genus in the Mediterranean Sea.

Materials and methods

Laboratory strains and growth conditions

All *Prochlorococcus* and *Synechococcus* strains used in this study were grown at 20 ± 1 °C in PCR-S11 medium (Rippka *et al.*, 2000). The origin of strains has been described elsewhere (Partensky *et al.*, 1993; Fuller *et al.*, 2003). *Synechocystis* PCC 6803 was cultivated at 25 °C in BG11 medium (Rippka *et al.*, 1979). Flow cytometry was used for cell counts and assaying culture growth (Marie *et al.*, 2005).

Sampling and mini-lysate preparation

Several stations along two transects in the Mediterranean Sea (from Gibraltar to a station located off the north-east coast of Libya and north-west to the French coast through the Tyrrhenian and Ligurian Seas) were sampled in September 1999 during the Prosope cruise aboard the R.V. *La Thalassa*. The cruise track and location of stations are shown in Fig. 1. Nutrients (nitrate and phosphate) were measured onshore as previously described (Raimbault *et al.*, 1990). At all stations and depths, samples were taken for flow cytometry (1.5 mL fixed with 0.1% glutaraldehyde for 15 min then frozen in liquid nitrogen) and for pigment analyses (2.8–5.6 L filtered on GF/F filters; Whatman International Ltd, Maidstone, UK) and quickly frozen in liquid nitrogen. For molecular diversity studies of *Prochlorococcus* population diversity, 2.5-L water samples were retrieved from 4 to 6 depths at stations 3, 5, MIO, 8, 9 and DYE, using Niskin bottles fitted on a Rosette sampler equipped with conductivity, temperature and depth (CTD) sensors. Seawater was prescreened over 200- μ M silk and 3- μ M Nuclepore filters (Whatman) and picoplankton were collected by filtration on to 47-mm diameter filters with a 0.2- μ m pore size (Pall Supor-200, Gelman Inc., Ann Arbor, MI). The filter was then transferred into a cryovial and quickly frozen in liquid nitrogen.

Mini-lysates were prepared for each sample to be used for PCR amplification. Filters were first cut into small pieces using a clean razor blade and resuspended in 392 μ L of lysis solution (200 mM dithiothreitol, 0.01% SDS) with four glass beads (~ 1 μ m average diameter). Tubes were vortexed for 1 min and submitted to three cycles of freezing in liquid nitrogen and thawing in a 55 °C water bath. Proteinase K was then added at a final concentration of 200 ng mL⁻¹ and

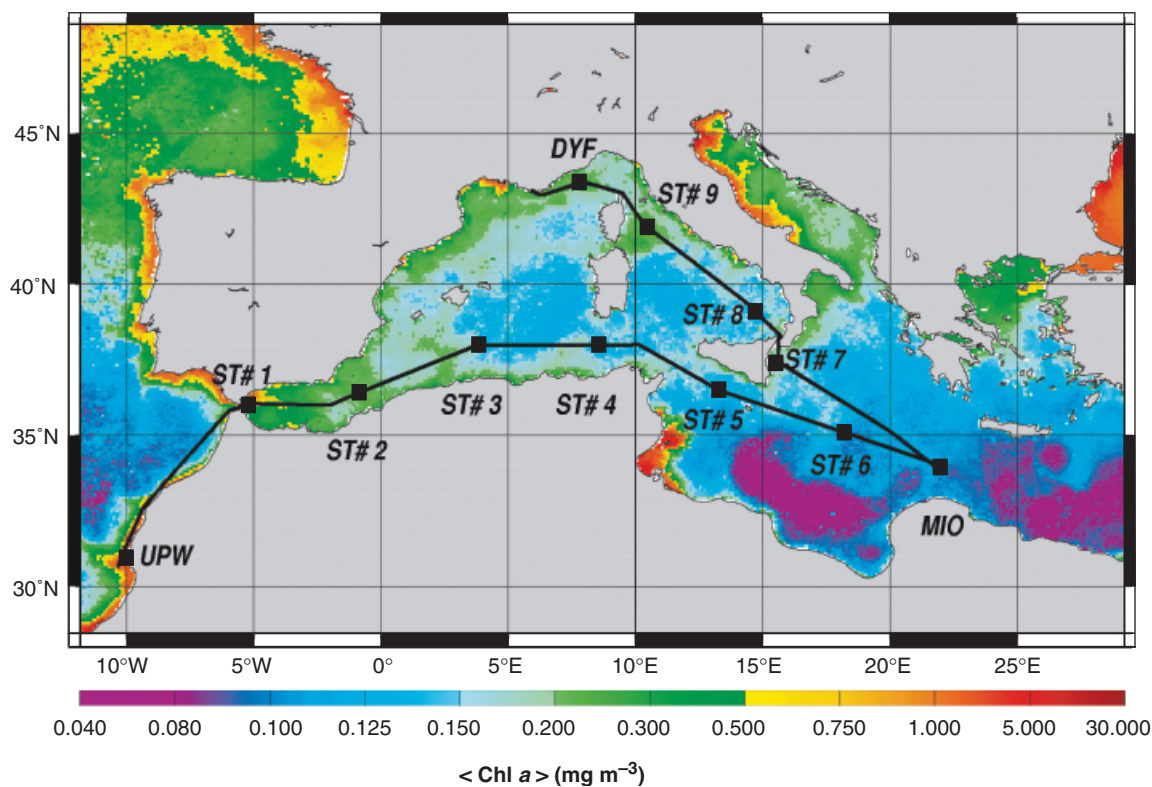


Fig. 1. Track of the Prosopé cruise superimposed on the composite SeaWiFS image of surface Chl *a* concentration for September 1999. ST#, station number.

the samples were incubated for 1 h at 55 °C. Mini-lysates were kept at – 80 °C until used for PCR.

Flow cytometric and HPLC pigment analysis

Flow cytometry samples were run on a FACSort flow cytometer (Becton Dickinson, San Jose, CA), as described by Marie *et al.* (1999). Pigment analyses were carried out using HPLC as described previously (Marie *et al.*, 2005).

Oligonucleotide probe hybridizations

Vertical profiles of different *Prochlorococcus* ecotypes were measured using dot-blot hybridization technology, as described previously (West & Scanlan, 1999). Briefly, probes designed against specific regions of the 16S rRNA gene of HL strains belonging to the HLI (S1PRO634R) and HLII (S2PRO634R) clades, of *P. marinus* SS120 (SARG634R), a LL strain which clusters apart in 16S rRNA gene trees, and of all LL strains but SS120 (DPRO634R) were end-labeled by $\gamma^{32}\text{P}$ prior to use. 16S rRNA gene amplicons obtained with the OXY107F and OXY1313R primers, biased to amplifying oxygenic phototrophs (West & Scanlan, 1999), were purified by the QIAquick kit (Qiagen, Hilden, Germany), quantified spectrophotometrically, then denatured. Thirty nanograms of these PCR products were blotted in triplicate on to nylon

membranes (Zetaprobe, Bio-Rad, Hercules, CA). Probes were prehybridized in Z-hyb buffer [1 mM EDTA, 0.25 mM Na_2HPO_4 (pH 7.2), 7% (w:v) SDS] for 30 min at 30 °C then hybridized overnight in 5 mL of fresh Z-hyb buffer containing labeled oligonucleotides. Membranes were then washed three times in $0.2 \times$ SSPE-0.15% SDS for 15 min at 30 °C, followed by a 10-min wash at the stringency temperature determined previously (West & Scanlan, 1999). Hybridization was quantified with a PHOSPHORIMAGER and the IMAGE QUANT software (Molecular Dynamics, GE Healthcare Bio-Sciences, Uppsala, Sweden). The relative level of hybridization (i.e. the probe signal as a proportion of the total DNA amplified with the OXY107F-OXY1313R primer pair) was then calculated by normalizing the individual *Prochlorococcus* probe signals against the signal for total amplified DNA as determined with the EUB338 probe, as reported previously (Gordon & Giovannoni, 1996; West & Scanlan, 1999).

Restriction fragment length polymorphism analysis

The diversity of natural *Prochlorococcus* populations was assessed by analyses of DNA band patterns obtained by RFLP analysis of *pcb* gene amplicons digested by HaeIII,

following a protocol modified from Garczarek *et al.* (2000). Briefly, *Prochlorococcus* spp. *pcb* genes were amplified by PCR (Ready-to-Go PCR beads kit, GE Healthcare Bio-Sciences, Uppsala, Sweden) using 4 µL of mini-lysate (see above) as template. After 5 min denaturation at 94 °C, amplification was performed by touchdown PCR: the annealing temperature was gradually reduced by 1 °C every cycle from 65 to 57 °C, followed by 25 cycles at 57 °C. PCR was terminated by a 10 min elongation step at 72 °C. Two sets of degenerate primers were optimized from an alignment comprising all available complete *pcb* sequences, including those of the HL-adapted *Prochlorococcus* strains MED4 (HLI), SB, GP2 and TAK9803–2 (HLII) and of the LL strains MIT9313 and SS120: the forward primer was *pcb*-1-fw-P1D (5'- ATGCARACYTAYGGDAAYCC-3'), while the reverse was either *pcb*-760-rev-P3 (5'-GCNGCNAT-DATNGCCATCCA-3') or *pcb*-988-rev-P4 (5'-ACNCKYT-TRAARTCRAANCC-3'). Amplicons from four reactions were pooled and purified with the Wizard kit (Promega, Madison, WI) before being digested with 10 U of restriction endonuclease HaeIII (Promega) for 2 h. Restriction fragments were resolved by gel electrophoresis on 2% Metaphor agarose gels (Seakem LE agarose, FMC Bioproducts, Rockland, ME). The size of RFLP bands was determined using the CROSS CHECKER software v. 2.91 (from J.B. Buntjer).

Gene sequencing and phylogenetic analysis

pcb genes were amplified using the *pcb*1-fw-P1D/*pcb*-988-rev-P4 primer set from *Prochlorococcus marinus* strain NATL1A and using the *pcb*1-fw-P1D/*pcb*-760-rev-P3 primer set from Prosopé stations (St.) 3 and 5 and then cloned using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA). For St. 3, 50 clones per depth (5, 55, 110) were screened by RFLP (HaeIII digestion) and those showing different banding patterns were sequenced bi-directionally using universal primers on an ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA). For St. 5, 29 samples from 5 m and 53 from 95 m were sequenced without previous RFLP screening. The derived Pcb protein sequences were automatically aligned using the CLUSTALW option (Thompson *et al.*, 1994) of the BIOEDIT freeware (Hall, 1999) and alignments were then manually refined. Regions of ambiguous alignment or hypervariability as well as primer regions were excluded from the analysis, leaving 239 amino acids residues. Potentially chimeric sequences, as assessed by the close match of their 5' region to one set of sequences and their 3' region to another set of sequences, were eliminated from the dataset. Phylogenetic analysis was performed using the Jones Taylor Thornton (JTT) model and the variability of substitution rates across sites and invariable sites were estimated by maximum likelihood using the phylogeny program PHYML v2.4.4 (Guindon & Gascuel, 2003). Boot-

strap analyses including 1000 replicate data sets were used to estimate the relative confidence in monophyletic groups. Phylogenetic trees were edited using the program MEGA3 v3.1 (Kumar *et al.*, 2004). Sequences of *pcb* genes obtained for environmental samples from the Mediterranean Sea have been deposited in the EMBL nucleotide sequence database under the following accession numbers: AM285349–AM285448. For *P. marinus* NATL1A *pcb* genes, EMBL accession numbers are reported in Table 1 (see below).

Results and discussion

Hydrology and distribution of *Prochlorococcus* cells along the Prosopé cruise transect

A Sea-viewing Wide Field-of-view Sensor (SeaWiFS) image in false color taken at the time of the Prosopé cruise (Fig. 1) showed a markedly decreasing gradient of surface Chl *a* concentrations from the mesotrophic St. 1, located in the Strait of Gibraltar, to St. MIO, situated in a highly oligotrophic area in the center of the Ionian Sea. This suggests that large changes in phytoplankton abundance and/or community structure occurred along this west–east (W–E) transect. Plots of the variations with depth of a number of physical and chemical characteristics along this transect (Fig. 2) showed the complexity of the studied zone in which several physical and chemical gradients were superimposed. Temperature and salinity transects (Figs. 2a and b) reflect the progression of low-salinity, low-temperature Atlantic waters (AW) penetrating through the Gibraltar strait and flowing along the African coast where they are progressively modified by interaction with the atmosphere (evaporation) and by mixing with Mediterranean waters (MW). Thus, a strong halocline is observed between St. 3 and 4, clearly identifying a partitioning of surface waters into AW (westwards) and typical high salinity MW (eastwards). Furthermore, there was a progressive W–E deepening of the nitracline (Fig. 2c) and phosphocline (Fig. 2d) down to 98 and 124 m depth, respectively, at St. MIO. St. 1–3 were located in the Algerian current, which is characterized by high spatial variability due to permanent hydrodynamic features such as up-wellings, fronts and eddies (Prieur & Sournia, 1994). Such a mesoscale variability could explain the anomalies in the general distribution of nutrients (Minas *et al.*, 1991; Raimbault *et al.*, 1993), as observed for nitrate at St. 1, which was surprisingly depleted in the upper layer, and at St. 3, where a secondary concentration minimum could be observed centered around 100 m, probably resulting from the recent down-welling of surface MW (Fig. 2c).

Figure 2e shows the depth distribution of *Prochlorococcus* cell concentration along the St. 1 – MIO transect. It was

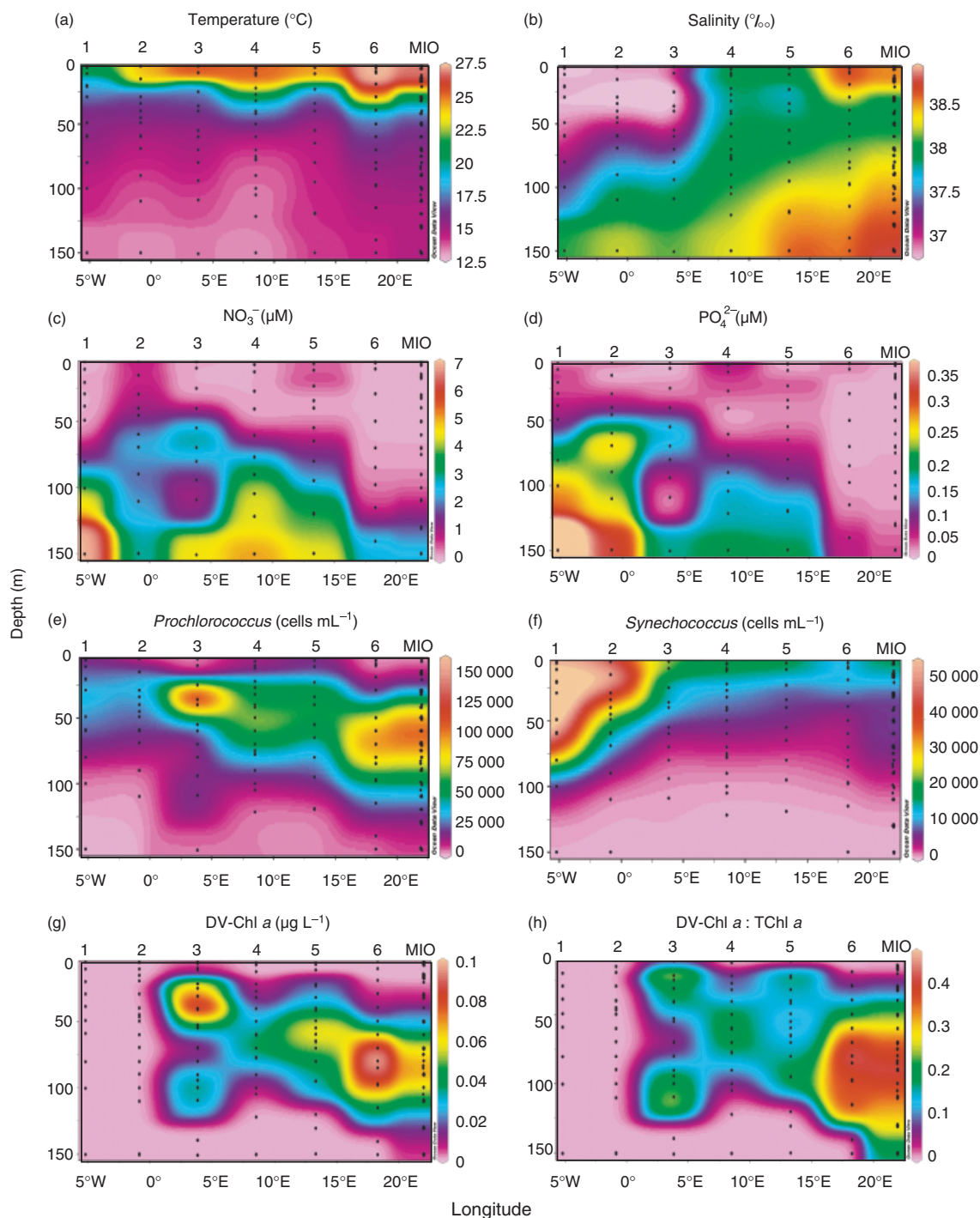


Fig. 2. Hydrological conditions and distribution of *Prochlorococcus* cell densities and divinyl-Chl *a* along the main W-E transect (station 1-MIO) of the Prosopé cruise. Vertical profiles of (a) temperature, (b) salinity, (c) nitrate, (d) phosphate, (e) *Prochlorococcus* cell concentrations, (f) *Synechococcus* cell concentrations, (g) divinyl-Chl *a* concentrations and (h) ratio of divinyl-Chl *a* to Chl *a*. Figures above the graph indicate station numbers and black dots correspond to sampled depths for each parameter.

markedly different from that of *Synechococcus* which were most abundant at the entrance of the Mediterranean Sea and restricted to the uppermost layer over the whole transect

(Fig. 2f). As expected, patterns of *Prochlorococcus* cell density globally followed the distribution of the *Prochlorococcus*-specific pigment DV-Chl *a*, with some exceptions

(Fig. 2g). Indeed, DV-Chl *a* could not be detected at St. 1 and 2, although *Prochlorococcus* cells were present and relatively abundant (up to *c.* 45 000 cells mL⁻¹ at 30 m). This is due to the fact that organisms containing monovinyl- (MV-) Chl *a* (i.e. normal Chl *a*), including many diatoms, largely dominated the Chl biomass at the entrance of the Mediterranean Sea (data not shown). At St. 3, there was both a major maximum of *Prochlorococcus* cell density in the subsurface (108 000 cells mL⁻¹ at 40 m) as well as a second minor maximum at depth (23 000 cells mL⁻¹ at 94 m; Fig. 2e), likely generated by the recent down-welling of surface MW (see above). This second maximum was even more conspicuous for the DV-Chl *a* profile (Fig. 2g) than for cell density because *Prochlorococcus* cells at depth had a higher DV-Chl *a* content per cell than cells located in the upper part of the water column (0.93 and 1.84 fg DV-Chl *a* per cell at 40 and 94 m, respectively). East of St. 3, there was a progressive deepening of the *Prochlorococcus* abundance peak down to *c.* 70 m at St. 6 and MIO. The DV-Chl *a* maximum, which generally coincided with the MV-Chl *a* maximum (Marie *et al.*, 2005), followed the same trend as the *Prochlorococcus* abundance maximum, but was located deeper (86 m in average at St. MIO). Indeed, in this part of the euphotic zone, the DV-Chl *a* content per cell increased more with depth than the cell density decreased (see a similar example in Partensky *et al.*, 1996). In the eastern part of the transect *Prochlorococcus* cells were apparently very scarce in the upper layer, although their very low red fluorescence rendered their precise enumeration by flow cytometry difficult. This scarcity (and/or very low Chl content per cell) in surface waters was confirmed by the low DV-Chl *a* concentrations found in these waters. This depth distribution with a strong deep maximum and low cell concentrations in the upper layer is reminiscent of the type of distributions observed in the north-east (NE) Atlantic at the oligotrophic, temperate station OFP in July (Olson *et al.*, 1990), but contrasts with those found year-round in most tropical or subtropical oligotrophic waters where *Prochlorococcus* cells can be found at high abundance up to the surface (Campbell & Vaulot, 1993; Partensky *et al.*, 1996; Campbell *et al.*, 1997). The contribution of *Prochlorococcus* to the total Chl *a* biomass can be assessed by the DV-Chl *a* to MV-Chl *a* ratio (Fig. 2h). It varied from negligible at the Strait of Gibraltar up to an average ~40% at the DV-Chl *a* maximum at St. MIO.

Distribution of *Prochlorococcus* ecotypes as analyzed by dot blot hybridization

Analyses of *Prochlorococcus* ecotype distribution with a dot blot hybridization approach and different specific probes were made at St. 3, 5, MIO and 9 (Fig. 3). In all cases, HL ecotypes dominated the upper mixed layer, with a

greater relative abundance at the thermocline than in surface waters, whereas LL ecotypes were dominant below the thermocline.

The HLI clade-specific probe systematically gave much higher hybridization signals than the HLII probe, suggesting that the HLI clade is the dominant HL group all over the Mediterranean Sea (at least in summer). It is worth noting that if these two clades generally co-occur, one of them generally largely overgrows the other one (West *et al.*, 2001; Johnson *et al.*, 2006). Which type is dominant is however variable between oceanic areas. Indeed, besides the Mediterranean Sea, the HLI clade is also predominant in temperate areas of the Atlantic Ocean, whereas the HLII clade has been found to predominate in inter-tropical Atlantic zones and in the Red Sea (West & Scanlan, 1999; West *et al.*, 2001; Johnson *et al.*, 2006). The differential distribution of these two ecotypes has also been associated with the degree of stratification of the water column, the HLII ecotype being prevalent in strongly stratified waters and a deep thermocline and HLI in waters with moderate stratification and mixed layer depth (Bouman *et al.*, 2006). Although genetically very close, HLI and HLII ecotypes have been shown to differ in their phosphate acquisition capabilities with the former type having a more complete set of P acquisition genes and displaying a more efficient alkaline phosphatase activity (Moore *et al.*, 2005). Thus the dominance of the HLI clade in the Mediterranean Sea may in part be due to a better ability to cope with low P levels (Krom *et al.*, 1991; Moutin *et al.*, 2002). HLII populations have indeed been shown to correlate negatively with P concentrations over larger spatial scales (Bouman *et al.*, 2006). However, the fact that HLII populations predominate in the Sargasso Sea, another oceanic region where P is thought to be a limiting nutrient (e.g. see Wu *et al.*, 2000) suggests the situation is more complicated. Indeed, Martiny *et al.* (2006) showed a high variability in P gene content among *Prochlorococcus* strains that was not congruent with rRNA gene phylogeny. Thus the acquisition by lateral transfer of a few key genes (such as alkaline phosphatase, phosphonate uptake genes, etc.) or their differential retention during the genome reduction process (Dufresne *et al.*, 2005) could rapidly change the spectrum of usable nutrient sources for a *Prochlorococcus* cell, independently from its phylogenetic position or ecological niche.

The *P. marinus* SS120-specific probe (SARG probe) generally gave a fairly low signal, often below 3.5% relative hybridization, whereas the second LL probe, recognizing all known LL *Prochlorococcus* strains but SS120, usually gave a much higher signal (Fig. 3). Therefore, SS120-like populations seem to be a minor component among LL clades in the Mediterranean Sea.

The sum of the relative hybridization signals from the different probes, which should theoretically cover most

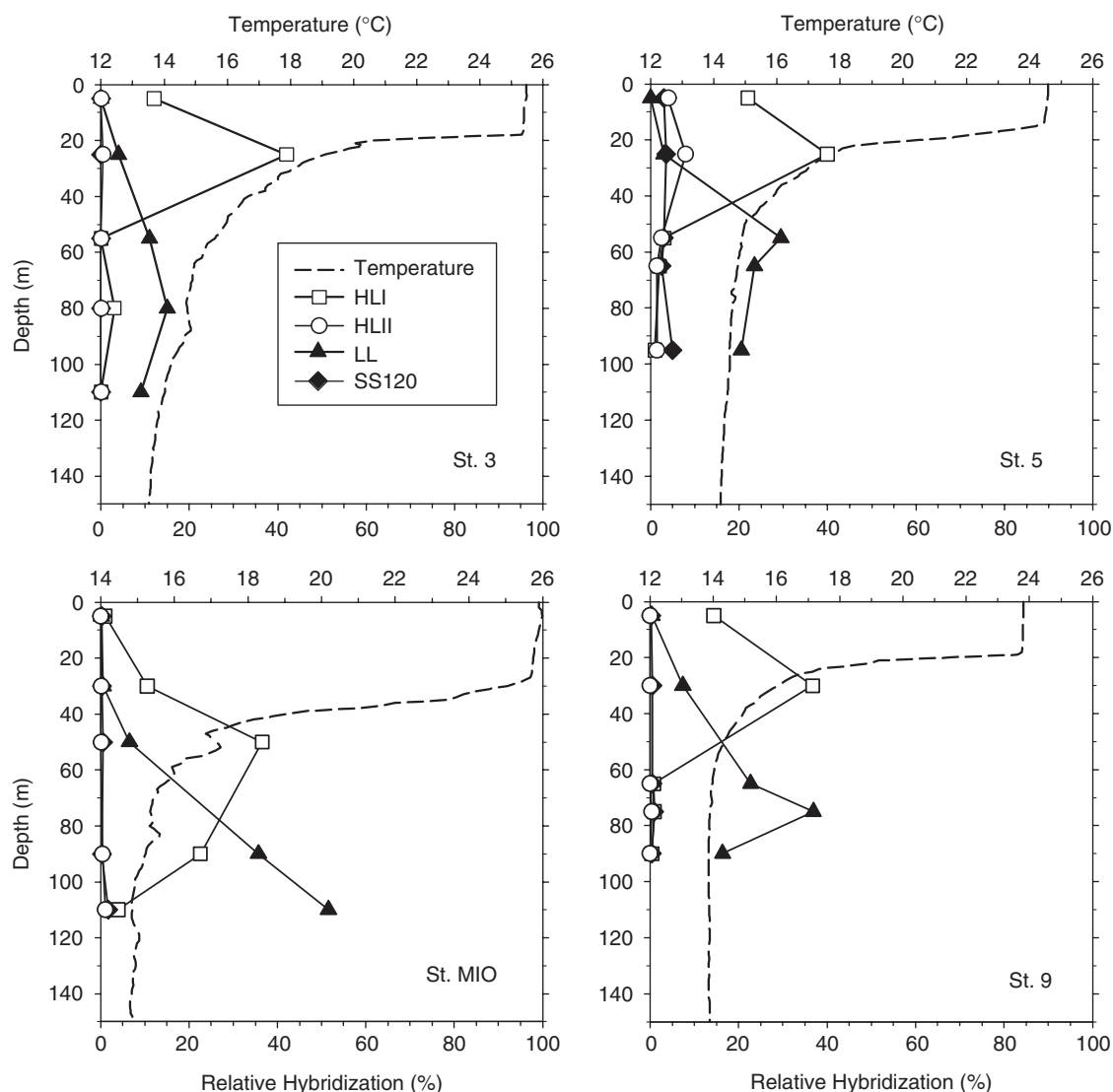


Fig. 3. Vertical distribution of *Prochlorococcus* ecotypes at four stations (St.) in the Mediterranean Sea analyzed by dot blot hybridization using 16S rRNA gene ecotype-specific oligonucleotides (HLI, high light type I; HLII, high light type II; LL, all LL ecotypes except SS120; SS120, Sargasso Sea type strain *P. marinus* SS120) showing relative hybridization of each ecotype and temperature down the water column. Relative hybridization represents the probe signal as a proportion of the total DNA amplified with the OXY107F-OXY1313R primer pair.

Prochlorococcus clades known to date (HLI, HLII and LL clades including SS120), was never higher than 59% (the value observed at St. MIO 90 m). In surface waters, the remainder of the hybridization signal can partially be attributed to *Synechococcus* cells which were abundant in these layers (see Fig. 2f). It is unlikely that *Prochlorococcus* HL cells were not recognized, as genetic diversity appears to be low among the numerous HL strains examined to-date from different ocean provinces (Rocap *et al.*, 2002). At depth, however, *Synechococcus* cells were not very abundant except at St. 1 and 2, and it is possible that unknown LL *Prochlorococcus* ecotypes were present.

Validation of the PCR-RFLP method using the *pcb* gene

In order to get a much finer resolution of *Prochlorococcus* ecotype diversity in the field, we applied a RFLP method that we previously used to characterize *Prochlorococcus* strains (Garczarek *et al.*, 2000). This method was shown to be able to discriminate the genetic signature of most tested strains, due both to the large variations in *pcb* gene sequences as well as variation in the number of bands. Before applying it to the natural environment, we performed a number of tests. Firstly, we tested the sensitivity of

our method using a range of cell concentrations (from 5×10^2 to 4×10^5 cells $^{-1}$) and different *Prochlorococcus* strains with two primer sets (pcb-1-fw-P1D/pcb-760-rev-P3 and pcb-1-fw-P1D/pcb-988-rev-P4; see 'Materials and methods'). These analyses revealed that the former set provided better results, as it allowed the amplification of *pcb* genes from most strains tested starting from as few as 1000 *Prochlorococcus* cells after mini-lysate preparation. However, we found that the PS I-specific *pcb* genes from *P. marinus* SS120 (i.e. *pcbC* and *pcbG*) as well as both *pcb* genes (*pcbA* and *pcbB*) from MIT9313 (Garczarek *et al.*, 2001; Bibby *et al.*, 2003), a high-G+C strain located at the base of the *Prochlorococcus* radiation (Rocap *et al.*, 2002, 2003), were more poorly amplified than other *pcb* types.

Secondly, because the *pcb* gene shows a strong homology to the *isiA* gene, which is induced under iron depletion in a number of cyanobacteria (see e.g. Bibby *et al.*, 2001; Boekema *et al.*, 2001), we tested whether the possible presence of *isiA* in natural samples might bias our analyses. The method was thus applied to the freshwater cyanobacterium *Synechocystis* PCC 6803 which is known to contain an *isiA* gene and to a number of marine *Synechococcus* spp. strains including WH7803, WH8102, RCC307 and CC9311. No fragment of the expected size (ca 770–800 bp) could be amplified from any of these strains, although the control (*Prochlorococcus* sp. MED4) gave a clear amplification product (data not shown). In fact, recent genome sequence information shows that no *isiA* gene is present in the three former *Synechococcus* spp. strains. In contrast, an homolog of *P. marinus* SS120 *pcbC*, potentially encoding an iron-induced, PS I-associated, light-harvesting protein similar to the IsiA/CP43' proteins found in a number of typical cyanobacteria (Bibby *et al.*, 2003), is present in several *Synechococcus* spp. strains from the Pacific Ocean, including CC9311, CC9902 and CC9605. Nevertheless, comparison of nucleotide sequence identity of these *Synechococcus pcbC/isiA*-like genes with our primers sequences showed that they could hardly be amplified, as confirmed by the absence of amplicon for CC9311.

Fine-scale ecotype diversity of *Prochlorococcus* in the Mediterranean Sea

A RFLP analysis of *pcb* amplicons was made at 4–6 depths at St. 3, 5, MIO, 8, 9 and DYF. Figures 4a–c illustrates the RFLP patterns obtained at three of these stations. Although band patterns cannot be taken as a truly quantitative assessment of the abundance of genes and/or ecotypes due to the possibility of preferential amplification of some *pcb* sequences over others (see next section), two major observations can be made from these gels. Firstly, for a given profile (e.g. St. 3), the pattern obtained at the uppermost depth (5 m) was very different from that obtained at 50–55 m or

below. The pattern observed at the bottom of the mixed layer (25–30 m) was often a combination of the surface one and that observed at the next sampled depth. Furthermore, although RFLP profiles from all depths below the thermocline shared nine bands (numbered 3–5, 13 and 15–19 on Fig. 4a), there was clearly a much larger number of minor bands in the ~0.26–~0.5 Kbp region in deep samples (e.g. St. 3, 80–110 m) than in intermediate ones (e.g. St 3, 25–55 m).

Secondly, comparison of samples from different Mediterranean Sea stations revealed a surprising similarity of band patterns at comparable depths throughout the sampled area, especially below the thermocline, patterns observed in the upper mixed layer being more variable. For instance, the pattern found at 55 m at St. 3 (Fig. 4a) was almost identical to that observed at the same depth at St. 5 (Fig. 4b) and at 30 m at St. 9 (Fig. 4c). Similarly, the pattern observed between 80 and 95 m at St. 3 was strikingly similar to that observed at 65 and 95 m at St. 5 and between 50 and 90 m at St. 9. These observations extend to all other sampled stations as well (data not shown). For comparison, application of our PCR-RFLP method to samples collected in September 2000 at different depths of St. A in the Gulf of Aqaba, Red Sea, gave RFLP patterns that were completely different from the Mediterranean Sea samples at all depths (Fig. 4d). Thus, the PCR-RFLP results obtained from the Mediterranean Sea clearly indicate a close relatedness (although not identity, see next section) between *pcb* genes amplified from very remote stations but similar light niches. This strongly suggests that there was a horizontal homogenization of ecotype diversity all over the sampled area.

The fact that the number of *pcb* genes varies drastically from one *Prochlorococcus* strain to another impacts on our ability to assess the number of ecotypes picked up by our PCR-RFLP method. However, by analyzing the relative abundance of the 19 bands discernible on our RFLP gels (see Fig. 4a) using the IMAGEQUANT software (Molecular Dynamics), we were able to distinguish a number of covarying *pcb* fragments which we assumed to belong either to a given *pcb* gene or to several *pcb* genes from a given ecotype, if the sum of fragments was higher than 0.8 Kbp (i.e. the average size of an undigested *pcb* amplicon obtained with the pcb-1-fw-P1D/pcb-760-rev-P3 primer set). Thus, this analysis provided a lower limit of the number of ecotypes present at each station.

All surface samples shared three major bands at ~0.80, ~0.66 and ~0.13 Kbp (bands 1, 2 and 17) that were also found in samples at the base of the thermocline but not below. Given the dot blot hybridization data (Fig. 3), it is reasonable to assume that these bands belonged to HL ecotype(s). The largest band (band 1) corresponds to undigested *pcb* gene(s) without HaeIII restriction sites.

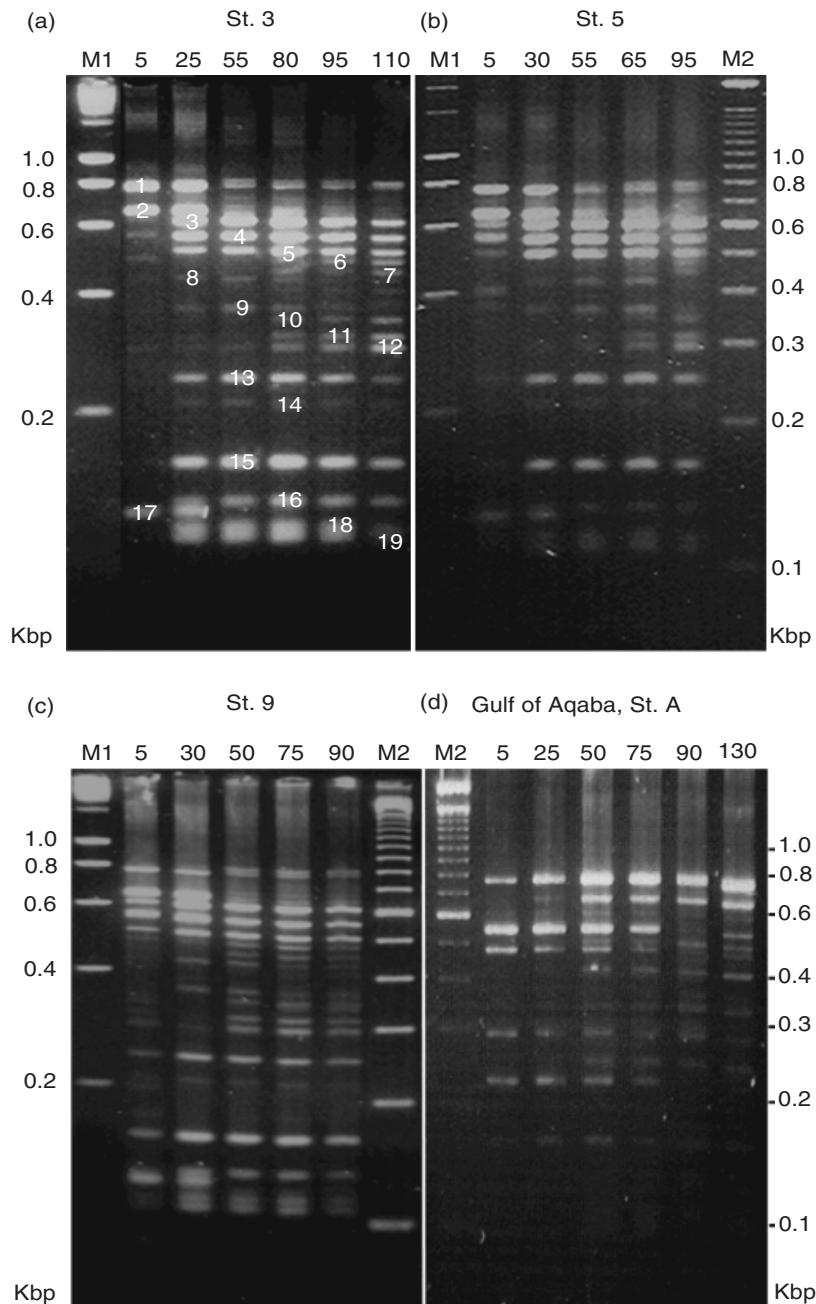


Fig. 4. RFLP patterns of PCR-amplified *pcb* genes at different depths (in meters) at three stations in the Mediterranean Sea (St. 3, 5 and 9) and at Station 'A' in the Gulf of Aqaba (29°28'N, 34°56'E), shown for comparison. M1: 100 kb scale marker; M2: 200 Kb scale marker.

Interestingly, bands 2 and 17 are similar in size to those obtained by digesting the sole *pcb* gene of the *P. marinus* HLI strain MED4, isolated from surface waters of the Mediterranean Sea (Table 1). The relatively simple pattern observed at 5 m at St. 3 suggests the presence of at least two co-occurring ecotypes with one *pcb* gene each (one of these ecotypes possibly being closely related to MED4) or, less likely (see next section), one major ecotype having two *pcb* genes (like

the HLII strain MIT9312). At most other stations, there were three additional intense bands at ~ 0.50 , ~ 0.56 and ~ 0.61 Kbp in the surface sample (bands 3, 4 and 5, which were already visible as faint bands at St. 3), suggesting the presence of at least three other abundant (and/or well amplified) *pcb* genes in these samples. However, bands of similar sizes were also found in deeper samples, down to 110 m.

At least three other groups of co-occurring bands, additional to the strictly HL ecotype (bands 1, 2, 17), can be observed in Fig. 4. Bands 8 (0.42 Kbp) and 9 (0.37 Kbp) which probably correspond to fragments of a single *pcb* gene (see next section), co-occurred over variable intermediate depth ranges, but were never found in the uppermost samples and only at trace amounts in the deepest samples. The co-occurring bands 6, 7, 10, 11 and 12 (0.47, 0.45, 0.35, 0.32 and 0.30 Kbp, respectively), which may represent fragments of multiple *pcb* genes from the same ecotype, were systematically restricted to deeper depths and were not found above 110 m at the very oligotrophic MIO station (not shown), in good agreement with the W-E deepening of the *Prochlorococcus* maximum (Fig. 2f). The multiplicity of bands at depth is consistent with the fact that *Prochlorococcus* LL ecotypes have several *pcb* genes (from two in MIT 9313 up to seven in NATL2A and MIT9211 and even eight in SS120). The last group is represented by the aforementioned bands 3–5 which seem to co-occur with the smaller bands 13, 14, 15 and 16 (0.25, 0.22, 0.16 and 0.13 Kbp, respectively) and 18–19 (0.12 and 0.10 Kbp, respectively; see Fig. 4c). They apparently belong to one (or several) ecotype(s) having multiple *pcb* genes and exhibiting a wide growth irradiance range. The possibility that a single *Prochlorococcus* ecotype could occupy niches with a wide range of light intensities is consistent with previous data from the Red Sea where HLII ecotypes were shown to occupy the entire water column (West *et al.*, 2001; Fuller *et al.*, 2005), as well as with data from the Atlantic Ocean (Johnson *et al.*, 2006). Indeed, the latter study showed that at high latitudes, such as that of the Mediterranean Sea, some ecotypes such as the LL clade including the NATL2A strain (the so-called eNATL2 ecotype) spanned most of the euphotic zone, whereas at lower latitudes it was restricted to deeper waters.

Comparative analyses of *pcb* gene clone libraries

In order to better interpret the restriction patterns obtained by PCR-RFLP and confirm the validity of the conclusions raised by this approach, we sequenced large fragments of *pcb* genes from natural *Prochlorococcus* populations collected at St. 3 and 5. PCR products from three depths at St. 3 (5, 55 and 110 m) were cloned, digested with HaeIII, and 1–2 clone(s) per restriction pattern were then selected for sequence analysis. For St. 5, we systematically sequenced (without previous selection based on restriction patterns) a large number of clones from two depths (5 and 95 m). This strategy aimed to (i) verify whether high sequence identity existed for clones retrieved from St. 3 and 5, as suggested by the PCR-RFLP analysis and (ii) correlate PCR-RFLP results with phylogenetic analyses (Fig. 5).

The 100 *pcb* sequences obtained, after omission of PCR chimeras, were all different, suggesting a large degree of

microdiversity among *Prochlorococcus* populations. However, a number of sequences were identical at the amino acid level, indicating that many of the nucleotide exchanges were synonymous, most of them being located at the third codon position. The predicted restriction patterns of these environmental sequences using HaeIII are reported in Table 2. In most cases, sequences belonging to a given restriction group (or to restriction groups having one or several restriction fragments in common and a similar total PCR product length) were very closely related. For instance, the paired DNA distances between the 12 environmental nucleotide sequences of restriction group RG26 ranged from 0.0096 to 0.0364, the lowest distance being found between St3_55_001 and St5_95m_026, i.e. samples collected at two different stations. The degree of genetic variation within the related restriction groups RG1 through RG3 was even lower, ranging from 0.0027 and 0.0318.

Among the 30 different restriction groups obtained for our Mediterranean Sea environmental sequences, only three gathered sequences from distantly related *pcb* genes, namely restriction groups RG4, RG10 and RG11. RG4 is not really a restriction group because it contains sequences lacking HaeIII restriction sites and a PCR product length of 776 bp, including the primer regions (see Table 2). Furthermore, in RG11, only one out of 18 sequences was an outlier (St3_110m_002), all others were closely related and clustered together in phylogenetic trees (see below). Thus, in most cases, restriction patterns can be considered as gene-specific (i.e. can differentiate *pcb* gene types) and ecotype-specific (i.e. can differentiate groups of genotypes belonging to the same ecotype, for a given *pcb* gene), and the degree of nucleotide variation within restriction groups was generally $\leq 4\%$. Consequently, our interpretation of PCR-RFLP patterns as indicating a large vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea appears to be validated. The comparison between RFLP and virtual restriction patterns allowed us to attribute some of the co-occurring *pcb* fragments in Fig. 4 to a given restriction group among our sequenced clones (Table 2), and therefore to a specific phylogenetic cluster (see below). For instance, the RFLP bands 2 and 17 obtained in surface samples (Fig. 4) correspond to restriction group RG5 (Table 2), including the *pcb* sequence from the HLI strain MED4 (Table 1). RFLP patterns from deeper samples are harder to interpret because of the multiplicity of bands and because many bands are shared between different restriction groups. Still, it is worth noting that restriction groups containing many Mediterranean Sea sequences (Table 2) have predicted fragment sizes similar to the major bands observed on RFLP gels (e.g. RG11 have fragment sizes similar to bands 3 and 4 in Fig. 4; RG26 to bands 4, 18 and 19; RG12 to bands 13, 15, 16 and 18; etc). More generally, using the list of all restriction bands from Table 2, it is possible to make a virtual RFLP

Table 1. Results from the virtual digestion by *Hae*III of the *pcb* gene region delineated by the *pcb*-1-fw-P1D/*pcb*-760-rev-P3 primer set (primer regions are included in the predicted fragment sizes) for a number of cultured *Prochlorococcus* strains

Strain	Ecotype	Locus tag (when issued from genome project)	Gene name	Predicted size of <i>Hae</i> III restriction fragments (bp)	Virtual restriction group	Genbank accession number	References for <i>pcb</i> sequence
MED4	HLI	–	<i>pcbA</i>	124; 652	RG5	NP_892745	(La Roche <i>et al.</i> , 1996; Rocap <i>et al.</i> , 2003)
SB	HLII	–	<i>pcbA</i>	292; 484	–	AAK69279	(Garczarek <i>et al.</i> , 2000)
GP2	HLII	–	<i>pcbA</i>	66; 124; 586	–	AAK69280	(Garczarek <i>et al.</i> , 2000)
TAK9803–2	HLII	–	<i>pcbA</i>	64; 159; 553	–	AAK69281	(Garczarek <i>et al.</i> , 2000)
MIT9312	HLII	PMT9312_0627	<i>pcbA</i>	776	RG4	YP_397124	Unpublished (Genbank)
MIT9313	LL	PMT9312_1271	<i>PcbC*</i>	65; 342; 369	–	YP_397766	
		PMT1046	<i>pcbA</i>	34; 36; 42; 89; 115; 120; 337	–	NP_894877	(Bibby <i>et al.</i> , 2003; Rocap <i>et al.</i> , 2003)
		Pro0783	<i>pcbA</i>	157; 159; 457	–	NP_875175	
SS120	LL	Pro1169	<i>pcbB</i>	74; 77; 145; 227; 253	–	NP_894329	
		Pro0885	<i>pcbC</i>	320; 456	–	NP_875277	
		Pro1167	<i>pcbD</i>	24; 74; 693	–	NP_875559	
		Pro1450	<i>pcbE</i>	50; 133; 159; 455	–	NP_875841	
		Pro1288	<i>pcbF</i>	211; 580	–	NP_875679	
		Pro0892	<i>pcbG</i>	110; 133; 159; 377	–	NP_875284	
		Pro1174	<i>pcbH</i>	74; 145; 260; 294	–	NP_875566	
		P9211_05132	<i>pcbA</i>	105; 292; 376	–	ZP_01005558	
		P9211_02952	<i>pcbB</i>	292; 478	RG9	ZP_01005122	
		P9211_03997	<i>pcbC*</i>	776	–	ZP_01005331	
MIT9211	LL	P9211_02962	<i>pcbD</i> [†]	24; 74; 180; 183; 330	–	ZP_01005124	Unpublished (Genbank)
		P9211_01582	<i>pcbE</i>	49; 122; 273; 365	–	ZP_01004848	
		P9211_01462	<i>pcbF</i>	791	–	ZP_01004824	
		P9211_02937	<i>pcbH</i>	24; 48; 85; 138; 478	–	ZP_01005119	
		PMN2A_0066	<i>pcbA</i>	376; 397	–	AAZ57558	
		PMN2A_0719	<i>pcbB1</i>	159; 611	RG11	AAZ58210	
		PMN2A_0723	<i>pcbB2</i>	107; 118; 133; 159; 253	RG12	AAZ58214	
NATL2A	LL	PMN2A_0215	<i>pcbC*</i>	61; 715	–	AAZ57707	Unpublished (Genbank)
		PMN2A_0718	<i>pcbD</i>	428; 363	–	AAZ58209	
		PMN2A_0722	<i>pcbE</i>	800	RG15	AAZ58213	
		PMN2A_0717	<i>pcbH</i>	107; 114; 552	RG26	AAZ58208	
		–	<i>pcbB1</i>	159; 227; 384	–	AM285346	
NATL1A	LL	–	<i>pcbB2</i>	107; 118; 133; 159; 253	RG12	AM2853467	
		–	<i>pcbD</i>	363; 428	–	AM2853465	This study
		–	<i>pcbE</i>	800	RG15	AM2853468	
Syn CC9902	Coastal	Syncc9902_1005	<i>pcbH</i>	107; 114; 552	RG26	AM2853464	
			<i>isiA</i>	17; 50; 84; 103; 107; 200; 201	–	YP_377013	Unpublished (Genbank)
Syn CC9605	Pelagic	Syncc9605_1590	<i>pcbC*</i>				
			<i>isiA</i>	14; 44; 64; 81; 93; 151; 153; 162	–	YP_381894	Unpublished (Genbank)
Syn CC9311	Coastal	Sync_2306	<i>pcbC*</i>				
			<i>isiA</i>	14; 21; 107; 120; 138; 162; 201	–	–	(Palenik <i>et al.</i> , 2006)

The sequences of five *pcb* genes from NATL1A were obtained in this study. Other sequences were retrieved from databanks (generally Genbank, unless specified otherwise). For LL-adapted strains except MIT9313, the last letter of the *pcb* gene names was attributed according to their homology (best hit in ExPasy) with previously characterized genes in *P. marinus* SS120 (Garczarek *et al.*, 2001; Bibby *et al.*, 2003). The sizes of restriction fragments were computed using the 'Restriction Digest Calculator' module of the 'Sequence manipulation Suite' (Stothard, 2000).

*Named *pcbD* in some databases, due to transitive annotation error.

[†]Named *pcbC* in some databases, due to transitive annotation error.

LL, low light-adapted; HL, high light-adapted.

Table 2. Results from virtual digestion by *HaeIII* of *pcb* genes amplified by the *pcb*-1-fw-P1D/*pcb*-760-rev-P3 primer set from natural samples collected in the Mediterranean Sea

Virtual restriction group	Computed size of <i>HaeIII</i> restriction fragments (bp)	Phylogenetic cluster	Sample names
RG1	87; 124; 565	HLI PcbA	St3_5m_001; St5_5m_003; St5_5m_004; St5_5m_010; St5_5m_015; St5_5m_021
RG2	211; 565		St5_5m_006; St5_5m_019
RG3	159; 617		St3_5m_003
RG4	776	HLIb PcbA	St3_5m_002; St5_5m_002; St5_5m_022 env_44268154 (MIT9312_pcbA)
		LL PcbC	St5_95m_041
RG5	124; 652	HLI PcbA	St5_5m_001; St5_5m_005; St5_5m_007; St5_5m_008; St5_5m_009; St5_5m_011; St5_5m_017; St5_5m_018; St5_5m_020; St5_5m_023; St5_5m_025; St5_5m_026; St5_5m_028; St5_5m_029; e44493307 (MED4_pcbA)
RG6	64; 159; 553	HLIIa PcbA	St5_5m_012; St5_5m_027
RG7	223; 553		env_43122726 (TAK98-3) env_44524025
RG8	214; 562	HLIb PcbA	env_44570220
RG9	292; 478	LL PcbB	St5_95m_043 (MIT9211_pcbB)
RG10	133; 159; 478		env. PcbE
RG11	159; 611	LL PcbB	St3_110m_011; St5_95m_045; St5_95m_047
		NATL-like PcbB1	St3_55m_004; St5_5m_024; St5_95m_011; St5_95m_014; St5_95m_015; St5_95m_022; St5_95m_028; St5_95m_031; St5_95m_034; St5_95m_037; St5_95m_040; St5_95m_046; St5_95m_048 (NATL2_pcbB1)
		env. PcbE	St3_110m_002
RG12	107; 118; 133; 159; 253	NATL-like PcbB2	St3_55m_003; St3_110m_006; St3_110m_012; St5_95m_036; (NATL2_pcbB2; NATL1_pcbB2)
RG13	118; 133; 159; 360		env_42902782
RG14	133; 159; 505	LL PcbE	St3_110m_014; St5_95m_023
		SS120-like PcbE	St5_95m_002; St5_95m_003
RG15	800	NATL-like PcbE	St5_95m_019 (NATL1_PcbE; NATL2_PcbE)
RG16	75; 342; 383	LL PcbE	St3_110m_007
RG17	133; 159; 508		St3_55m_002; St3_110m_003; St3_110m_010; St5_95m_008; St5_95m_010; St5_95m_021; St5_95m_025
RG18	133; 159; 251 254	SS120-like PcbE	St5_95m_009
RG19	36; 256; 478	env. PcbE	St5_95m_033

Table 2. Continued

Virtual restriction group	Computed size of Haelll restriction fragments (bp)	Phylogenetic cluster	Sample names
RG20	797	LL PcbE	St5_95m_024
RG21	50; 292; 455		St5_95m_020
RG22	86; 292; 419		St5_95m_013
RG23	73; 74; 170; 453		St5_95m_035
RG18	74; 354; 363	LL PcbD	env_43957883
RG19	133; 161; 180; 317		St5_95m_030
RG20	74; 133; 161; 180; 243		St3_110m_015
RG21	791		St5_95m_007
RG22	9; 24; 72; 74; 112; 201; 299	env. PcbD (LL PcbD)	St3_110m_009
RG23	9; 24; 72; 74; 612		St3_110m_005; St3_110m_008
RG24	9; 24; 72; 74; 299; 313		St3_110m_016; St5_95m_017; St5_95m_042
RG25	24; 74; 81; 157; 180; 275		St3_110m_013
RG26	107; 114; 552	NATL-like PcbH	St3_55m_001; St5_5m_013; St5_5m_014; St5_95m_001; St5_95m_004; St5_95m_006; St5_95m_012; St5_95m_016; St5_95m_018; St5_95m_026; St5_95m_032; St5_95m_039; (NATL1_pcbH; NATL2_pcbH)
RG27	221; 552		St5_95m_038
RG28	107; 666		env_42867948
RG29	65; 92; 107; 518	unknown env. LL Pcb	St5_95m_044
RG30	321; 449	LL PcbF	St3_110m_001

For comparison, *pcb*-like gene sequences from the Sargasso Sea (Venter *et al.*, 2004) were also included in the restriction analysis (coded as env_XXXXXXXX). Sizes of restriction fragments have been computed as explained in the legend of Table 1. To each restriction group (RG) corresponds one (and sometimes several) phylogenetic clusters on the maximum likelihood tree (see Fig. 5). When a *pcb* gene from a cultured *Prochlorococcus* strain has a similar virtual restriction group as a *pcb* gene from the field, the strain name is reminded between brackets.

gel (not shown) which closely resembles our actual RFLP gels (Fig. 4).

Pcb sequences from *P. marinus* NATL1A, environmental sequences retrieved from the Mediterranean Sea, as well as all publicly available Pcb sequences from HL and LL-adapted *Prochlorococcus* strains (listed in Table 1) and environmental sequences obtained by Venter *et al.* (2004) (Table 2) were used to construct a phylogenetic tree (Fig. 5). Pcb sequences from NATL1A were closely related to their homologues in the entirely sequenced *P. marinus* strain NATL2A (both strains were isolated in 1990 by FP during the same cruise of the R.V *Hudson*, but at different stations of the North Atlantic). At the nucleotide level, interstrain DNA distances ranged from 0.0039 for *pcbE* to 0.0175 for *pcbD*. We did not obtain any homologues of the *pcbA* and *pcbC* genes found in NATL2A, suggesting that those genes were poorly amplified from NATL1A with our primer set.

This is confirmed by the fact that many environmental sequences were closely related to NATL2A (or NATL1A) PcbB1, PcbB2, PcbD, PcbE and PcbH (Fig. 5), but only one environmental sequence was related to NATL2A PcbC (i.e. sample St5_95m_041) and none to NATL2A PcbA.

Protein sequences related to Pcb were also found by BLASTP analysis (Altschul *et al.*, 1997) of the genome databases of *Synechococcus* spp. CC9902, CC9605 and CC9311 and were used to root the maximum likelihood Pcb tree shown in Fig. 5. The presence of these homologues of the PS I-associated, iron stress-induced IsiA protein of freshwater cyanobacteria (Bibby *et al.*, 2001; Boekema *et al.*, 2001) in some (but not all) marine *Synechococcus* strains suggests that *isiA* must have been present in their common ancestor with *Prochlorococcus* and was subsequently lost in some *Synechococcus* lineages. In contrast, in *Prochlorococcus*, an early duplication of this ancestral gene must have occurred. One

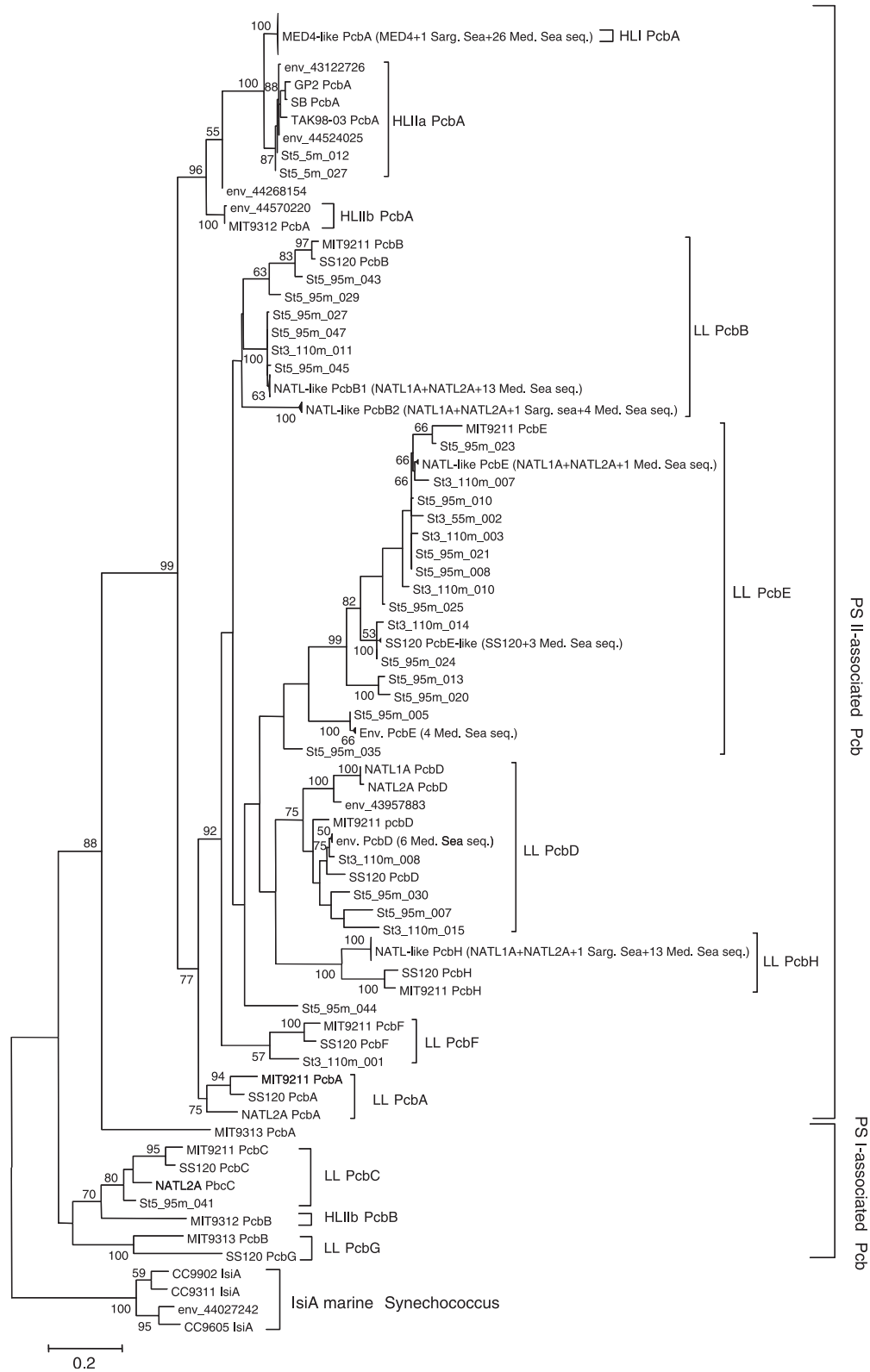


Fig. 5. Maximum likelihood tree of Pcb amino acid sequences (all gaps were excluded from the alignment) obtained with the PHYLX software. Only bootstrap values > 50% for 1000 replicates are shown. Details on sequences used in this analysis are listed in Tables 1 and 2. For readability, clusters containing a large number of closely related sequences are shown as vertical lines or triangles (depending on the variability among sequences) and the content of these clusters is indicated between parentheses. Med., Mediterranean; Sarg., Sargasso; seq., sequences.

of the copies continued to encode an iron-induced, PS I-associated protein form [e.g. PcbB in MIT9313 and PcbC in SS120 (Bibby *et al.*, 2003)], whereas the other differentiated to encode a PS II-associated Pcb protein. The gene encoding the PS I-associated Pcb was subsequently lost in some high-light adapted lineages, such as the HLI cluster containing MED4. This ancestral differentiation into PS I- and PS II-associated Pcb sequences has led to the two major Pcb branches of the phylogenetic tree shown in Fig. 5. The reason why a further multiplication of PS II-related genes occurred in some LL lineages (reaching up to six copies in NATL2A and SS120 strains) is possibly related to their adaptation to the LL niche. Interestingly, this gene duplication did not occur in the 'primitive' LL lineage containing MIT9313.

Twenty-six Mediterranean Sea sequences, all retrieved from surface (5 m) samples, appeared to be closely related to HLI *Prochlorococcus* strain MED4 and only two (St5_5m_012 and St5_5m_027) to the HLIIa strains SB, GP2 and TAK98-03. This seems to confirm that the major HL ecotype in the Mediterranean Sea is HLI and not HLII, in accordance with the Dot Blot Hybridization analysis (see above). Surprisingly, the MIT9312 strain (HLIIB clade in Fig. 5) had a fairly divergent PcbA sequence compared with other HLII strains (so-called HLIIa clade). One environmental sequence related to the HLIIB clade was found in the Sargasso Sea data set, but none were found in the Mediterranean Sea set. All other environmental sequences could be unambiguously identified as Pcb sequences from LL ecotypes and generally clustered with one or several previously sequenced Pcb sequences, allowing us to identify and name most of the clusters (Fig. 5). All environmental sequences but one (St5_95m_041) were retrieved in the branch of PS II-associated Pcb proteins, which confirms that our primer set poorly amplified *pcb* encoding PS I-related antenna proteins. Interestingly, there was very little diversity among our clones for the PcbB- and PcbH-like proteins and many environmental sequences within these clusters were more closely related to sequences from NATL strains than to SS120 and MIT9211 strains. Furthermore, we retrieved Mediterranean Sea sequences related to both PcbB forms found in NATL strains (so-called PcbB1 and PcbB2 in Table 1 and Fig. 5), whereas the two other LL strains had only one such protein form. Comparatively, there was a much higher diversity within the PcbD-like and PcbE-like protein clusters, and the environmental sequences found in these clusters were not necessarily closely related to corresponding NATL sequences. In particular, all the Mediterranean Sea PcbD-like proteins clustered with MIT9211 and SS120 PcbD and not with NATL PcbD. Thus, one of the dominant LL ecotypes present during the Prosopé cruise apparently possessed several *pcb* genes related to the North Atlantic strains NATL1A and NATL2A, but its other

pcb genes might have diverged more or have been acquired by lateral transfer from other ecotypes. However, given the PCR bias, one cannot exclude that these were simply less well amplified than co-occurring *pcb* genes from other ecotypes and were outnumbered during the amplification process.

Conclusions

Very large changes in hydrological (temperature, salinity, etc.) and chemical (nutrients) parameters were monitored along the main W-E transect of the Prosopé cruise (St. 1-MIO) in the Mediterranean Sea at the end of the summer. However, all the molecular analyses used here to study *Prochlorococcus* diversity conclusively demonstrated that populations exhibited much larger phylogenetic variability along depth profiles, where strong and opposing gradients of light and nutrient availability were recorded, than across the W-E transect. We obtained strong evidence from PCR-RFLP data on *pcb* genes that at least four different *Prochlorococcus* ecotypes existed along the Prosopé transect, each occupying different (though overlapping) niches of the photic zone. Thus, one ecotype (HLI) was confined to the upper mixed layer, a second one to intermediate depths, and a third ecotype to the bottom of the photic zone. A fourth ecotype, apparently related to NATL strains based on digestion patterns of *pcb* clones, has a seemingly much larger depth distribution. This conclusion is consistent with recent data using quantitative real time RT-PCR which also showed that the main 16S rRNA gene clades identified so far in culture correspond to different *Prochlorococcus* ecotypes with clearly distinct depth distributions in the field (Ahlgren *et al.*, 2006; Johnson *et al.*, 2006; Zinser *et al.*, 2006). In contrast, we have evidence for a horizontal homogeneity of these *Prochlorococcus* ecotypes, despite a large microdiversity at the genotype level, mainly attributable to synonymous nucleotide exchange. The main notable change when comparing different stations over the transect was a vertical spreading of the ecotype distribution, evidenced by the deepening of bands attributed to the 'deep' ecotype at St. MIO (not shown). This is consistent with the W-E deepening of nutriclines (Fig. 2c and d) and increase in water clearness (the 1% surface irradiance depth, a parameter defining the euphotic layer Z_{eu} , was at 70 m at St. 3 and 5 and at 105 m at St. MIO). However, the clear separation of modified Atlantic waters and typical Mediterranean waters occurring between St. 3 and 4 did not translate into any notable alteration in the vertical genetic structure of *Prochlorococcus* populations. This observation strongly suggests that for *Prochlorococcus* cells the most significant parameter is the ambient photon flux as it narrows their depth distribution to a restricted part of the photic layer (depending on the ecotype). In a given light niche, cells appear to be

relatively insensitive to variations in other physicochemical factors as long as the range of variation remains suitable for them to survive. Nevertheless, as Martiny *et al.* (2006) noticed for phosphorus metabolism genes, different genotypes of the same ecotype (i.e. inhabiting the same light niche) can possess distinct nutrient gene complements providing them with different nutrient assimilation capacities, independently from their phylogenetic position. It is possible that such differences occurred within a given ecotype along the transect, but our molecular methods based on 16S rRNA and *pcb* genes would not have detected them.

When applied to the field, the PCR-RFLP method allowed us to differentiate ecotypes (or some intermediate level in-between ecotypes and genotypes). One of the main advantages of this method was that it smoothed out slight differences in *pcb* gene sequences. Without this smoothing approach, we would likely not have realized just how genetically closely related ecotypes inhabiting similar light niches at different Mediterranean Sea stations really were. Interestingly, this method does not distinguish NATL1A from NATL2A, two strains isolated from two distinct stations but at similar depths of the NW Atlantic, and global comparison of these two strains (recently both totally sequenced) should allow us to see what level of diversity exists between these two different genotypes of the same ecotype.

Acknowledgements

We wish to thank the captain and crew of the R.V. 'L'Atalante'. Dominique Tailliez is warmly thanked for collecting and correcting CTD data. We also thank Brian Palenik for providing us some dried *Synechococcus* sp. CC9311 cells for PCR-RFLP tests. The work was supported by the European Community program MARGENES (QLRT 2001–01226), the national PROOF programs PROSOPE, BIOSOPE and UVECO (CNRS-INSU), as well as the regional program IMPALA (Région Bretagne).

References

- Ahlgren NA, Rocap G & Chisholm SW (2006) Measurement of *Prochlorococcus* ecotypes using real-time PCR reveals different abundances of genotypes with similar light physiologies. *Environ Microbiol* **8**: 441.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Bibby TS, Nield J & Barber J (2001) Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* **412**: 743–745.
- Bibby TS, Mary I, Nield J, Partensky F & Barber J (2003) Low-light-adapted *Prochlorococcus* species possess specific antennae for each photosystem. *Nature* **424**: 1051–1054.
- Boekema EJ, Hifney A, Yakushevskaya AE *et al.* (2001) A giant chlorophyll-protein complex induced by iron deficiency in cyanobacteria. *Nature* **412**: 745–748.
- Bouman HA, Ulloa O, Scanlan DJ *et al.* (2006) Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**: 918–921.
- Buck KR, Chavez FP & Campbell L (1996) Basin-wide distributions of living carbon components and the inverted trophic pyramid of the central gyre of the North Atlantic Ocean, summer 1993. *Aquat Microb Ecol* **10**: 283–298.
- Campbell L & Vaulot D (1993) Photosynthetic picoplankton community structure in the subtropical North Pacific Ocean near Hawaii (station ALOHA). *Deep-Sea Res* **40**: 2043–2060.
- Campbell L, Nolla HA & Vaulot D (1994) The importance of *Prochlorococcus* to community structure in the central North Pacific Ocean. *Limnol Oceanogr* **39**: 954–961.
- Campbell L, Liu HB, Nolla HA & Vaulot D (1997) Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991–1994 ENSO event. *Deep-Sea Res I* **44**: 167–192.
- Dufresne A, Salanoubat M, Partensky F *et al.* (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci USA* **100**: 10020–10025.
- Dufresne A, Garczarek L & Partensky F (2005) Accelerated evolution associated with genome reduction in a free-living prokaryote. *Genome Biol* **6**: 1–10.
- Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF & Scanlan DJ (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* **69**: 2430–2443.
- Fuller NJ, West NJ, Marie D, Yallop M, Rivlin T, Post AF & Scanlan DJ (2005) Dynamics of community structure and phosphate status of picocyanobacterial populations in the Gulf of Aqaba, Red Sea. *Limnol Oceanogr* **50**: 363–375.
- Garcia-Pichel F, Belnap J, Neuer S & Schanz F (2003) Estimates of cyanobacterial biomass and its distribution. *Algol Studies* **109**: 213–228.
- Garczarek L, Hess WR, Holtzendorff J, van der Staay GW & Partensky F (2000) Multiplication of antenna genes as a major adaptation to low light in a marine prokaryote. *Proc Natl Acad Sci USA* **97**: 4098–4101.
- Garczarek L, van der Staay GWM, Hess WR, Le Gall F & Partensky F (2001) Expression and phylogeny of the multiple antenna genes of the low-light-adapted strain *Prochlorococcus marinus* SS120 (Oxyphotobacteria). *Plant Mol Biol* **46**: 683–693.
- Gordon DA & Giovannoni SJ (1996) Detection of stratified microbial populations related to *Chlorobium* and *Fibrobacter* species in the Atlantic and Pacific Oceans. *Appl Environ Microbiol* **62**: 1171–1177.

- Guindon S & Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *System Biol* **52**: 696–704.
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.
- Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EM & Chisholm SW (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**: 1737–1740.
- Krom MD, Kress N, Brenner S & Gordon LI (1991) Phosphorus limitation of primary production in the eastern Mediterranean Sea. *Limnol Oceanogr* **36**: 424–432.
- Kumar S, Tamura K & Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**: 150–163.
- La Roche J, Van der Staay GWM, Partensky F *et al.* (1996) Independent evolution of the prochlorophyte and green plant chlorophyll *a/b* light-harvesting proteins. *Proc Natl Acad Sci USA* **93**: 15244–15248.
- Liu HB, Suzuki K, Minami C, Saino T & Watanabe M (2002) Picoplankton community structure in the subarctic Pacific Ocean and the Bering Sea during summer 1999. *Mar Ecol Progr Ser* **237**: 1–14.
- Marie D, Brussaard C, Partensky F & Vault D (1999) Flow cytometric analysis of phytoplankton, bacteria and viruses. *Current Protocols in Cytometry* (Robinson JP, ed), pp. 11.11.11–11.11.15. John Wiley and sons, Somerset, USA.
- Marie D, Zhu F, Balagué V, Ras J & Vault D (2005) Eukaryotic picoplankton communities of the Mediterranean Sea in summer assessed by molecular approaches (DGGE, TTGE, QPCR). *FEMS Microbiol Ecol* **55**: 403–415.
- Martiny AC, Coleman ML & Chisholm SW (2006) Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci USA* **103**: 12552–12557.
- Minas H, Coste B, Le Corre P, Minas M & Raimbault P (1991) Biological and geochemical signatures associated with the water circulation through the strait of Gibraltar and in the Alboran sea. *J Geophys Res* **96**: 8755–8771.
- Moore LR & Chisholm SW (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates. *Limnol Oceanogr* **44**: 628–638.
- Moore LR, Rocap G & Chisholm SW (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**: 464–467.
- Moore LR, Ostrowski M, Scanlan DJ, Feren K & Sweetsir T (2005) Ecotypic variation in phosphorus acquisition mechanisms within marine picocyanobacteria. *Aquat Microb Ecol* **39**: 257–269.
- Moutin T, Thingstad TF, Van Wambeke F, Marie D, Slawyk G, Raimbault P & Claustre H (2002) Does competition for nanomolar phosphate supply explain the predominance of the cyanobacterium *Synechococcus*? *Limnol Oceanogr* **47**: 1562–1567.
- Olson RJ, Zettler ER, Altabet MA, Dusenberry JA & Chisholm SW (1990) Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. *Deep-Sea Res* **37**: 1033–1051.
- Palenik B, Ren QH, Dupont CL, *et al.* (2006) Genome sequence of *Synechococcus* C39311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci USA* **103**: 13555–13559.
- Partensky F & Garczarek L (2003) The photosynthetic apparatus of chlorophyll *b*- and *d*-containing Oxychlorobacteria. *Photosynthesis in Algae, Vol. 14* (Larkum AWD, Douglas SE & Raven JA, eds), pp. 29–62. Kluwer Academic Publishers, Dordrecht, the Netherlands.
- Partensky F, Hoepffner N, Li WKW, Ulloa O & Vault D (1993) Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated from the North Atlantic and the Mediterranean Sea. *Plant Physiol* **101**: 285–296.
- Partensky F, Blanchot J, Lantoine F, Neveux J & Marie D (1996) Vertical structure of picophytoplankton at different trophic sites of the tropical northeastern Atlantic Ocean. *Deep-Sea Res I* **43**: 1191–1213.
- Partensky F, Blanchot J & Vault D (1999a) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Marine Cyanobacteria, Vol. 19*. (Charpy L & Larkum AWD, eds), pp. 457–475. Bulletin de l'Institut Oceanographique, Monaco.
- Partensky F, Hess WR & Vault D (1999b) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63**: 106–127.
- Prieur L & Sournia A (1994) 'Almofront-1' (April–May 1991): an interdisciplinary study of the Algeria-Oran geostrophic front, SW Mediterranean Sea. *J Mar Syst* **5**: 187–203.
- Raimbault P, Slawyk G, Coste B & Fry J (1990) Feasibility of using an automated colorimetric procedure for the determination of seawater nitrate in the 0 to 100 nM range: examples from field and culture. *Mar Biol* **104**: 347–351.
- Raimbault P, Coste B, Boulhadid M & Boudjellal B (1993) Origin of high phytoplankton concentration in deep chlorophyll maximum (DCM) in a frontal region of the Southwestern Mediterranean Sea (Algerian Current). *Deep Sea-Res* **40**: 791–804.
- Rippka R, Deruelles J, B WJ, Herdmann M & Stanier RY (1979) Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111**: 1–61.
- Rippka R, Coursin T, Hess W *et al.* (2000) *Prochlorococcus marinus* Chisholm *et al.* 1992 subsp. *pastoris* subsp. nov. strain PCC 9511, the first axenic chlorophyll *a*₂/*b*₂-containing cyanobacterium (Oxyphotobacteria). *Intl J Syst Evol Microbiol* **50**: 1833–1847.
- Rocap G, Distel DL, Waterbury JB & Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68**: 1180–1191.

- Rocap G, Larimer FW, Lamerdin J *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.
- Six C, Thomas JC, Thion L, Lemoine Y, Zal F & Partensky F (2005) Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium *Synechococcus* sp strain WH8102. *J Bacteriol* **187**: 1685–1694.
- Stothard P (2000) The sequence manipulation suite: javascript programs for analyzing and formatting protein and DNA sequences. *Biotechniques* **28**: 1102–1104.
- Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Venter CJ, Remington K, Heidelberg JF *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **307**: 66–74.
- West NJ & Scanlan DJ (1999) Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern north Atlantic ocean. *Appl Environ Microbiol* **65**: 2585–2591.
- West NJ, Schonhuber WA, Fuller NJ, Amann RI, Rippka R, Post AF & Scanlan DJ (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. *Microbiol Sgm* **147**: 1731–1744.
- Wu JF, Sunda W, Boyle EA & Karl DM (2000) Phosphate depletion in the western North Atlantic Ocean. *Science* **289**: 759–762.
- Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, Scanlan DJ & Chisholm SW (2006) *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl Environ Microbiol* **72**: 723–732.