

## ORIGINAL ARTICLE

# A global perspective on marine photosynthetic picoeukaryote community structure

Amy R Kirkham<sup>1,5,7</sup>, Cécile Lepère<sup>1,6,7</sup>, Ludwig E Jardillier<sup>2</sup>, Fabrice Not<sup>3</sup>, Heather Bouman<sup>4</sup>, Andrew Mead<sup>1</sup> and David J Scanlan<sup>1</sup>

<sup>1</sup>*School of Life Sciences, University of Warwick, Coventry, UK;* <sup>2</sup>*Unité d'Ecologie, Systématique et Evolution, UMR CNRS 8079 Université, Paris-Sud, Paris, France;* <sup>3</sup>*CNRS and Université Pierre et Marie Curie, UMR 7144, Station Biologique de Roscoff, Roscoff, France and* <sup>4</sup>*Department of Earth Sciences, University of Oxford, Oxford, UK*

**A central goal in ecology is to understand the factors affecting the temporal dynamics and spatial distribution of microorganisms and the underlying processes causing differences in community structure and composition. However, little is known in this respect for photosynthetic picoeukaryotes (PPEs), algae that are now recognised as major players in marine CO<sub>2</sub> fixation. Here, we analysed dot blot hybridisation and cloning–sequencing data, using the plastid-encoded 16S rRNA gene, from seven research cruises that encompassed all four ocean biomes. We provide insights into global abundance,  $\alpha$ - and  $\beta$ -diversity distribution and the environmental factors shaping PPE community structure and composition. At the class level, the most commonly encountered PPEs were Prymnesiophyceae and Chrysophyceae. These taxa displayed complementary distribution patterns, with peak abundances of Prymnesiophyceae and Chrysophyceae in waters of high (25:1) or low (12:1) nitrogen:phosphorus (N:P) ratio, respectively. Significant differences in phylogenetic composition of PPEs were demonstrated for higher taxonomic levels between ocean basins, using Unifrac analyses of clone library sequence data. Differences in composition were generally greater between basins (interbasins) than within a basin (intra-basin). These differences were primarily linked to taxonomic variation in the composition of Prymnesiophyceae and Prasinophyceae whereas Chrysophyceae were phylogenetically similar in all libraries. These data provide better knowledge of PPE community structure across the world ocean and are crucial in assessing their evolution and contribution to CO<sub>2</sub> fixation, especially in the context of global climate change.**

*The ISME Journal* (2013) 7, 922–936; doi:10.1038/ismej.2012.166; published online 31 January 2013

**Subject Category:** microbial population and community ecology

**Keywords:** photosynthetic picoeukaryotes; community structure; prymnesiophytes; chrysophytes; global scale

## Introduction

Understanding the mechanisms that generate and maintain biodiversity, and more particularly the spatial distributions of taxa, is a key objective in ecology. This is essential to predicting ecosystem responses to future environmental change. However, most of our knowledge about how biodiversity varies across large spatial scales comes from research on macroorganism biogeography, with only

relatively few recent studies for microorganisms (for example, see Foissner, 2006; Fuhrman *et al.*, 2008; Caron, 2009; Nolte *et al.*, 2010). Opposing views have claimed that, on one hand, free-living microbial taxa present a cosmopolitan distribution (Finlay and Fenchel, 2004; Finlay *et al.*, 2006; Pither, 2007), with particular implications for organisms in the smallest size classes (for example, <0.5  $\mu$ m; Yang *et al.*, 2010), and, on the other, are composed primarily of species that have limited geographical distributions (Papke and Ward, 2004; Telford *et al.*, 2006; H rnstr m *et al.*, 2011).

During the past decade, molecular techniques have begun to provide insights into the structure of small (<3  $\mu$ m diameter), unicellular photosynthetic picoeukaryote (PPE) communities in various marine environments (reviewed by Vaultot *et al.*, 2008). In particular, cloning and sequencing of both nuclear (for example, see Moon van der Staay *et al.*, 2001; Not *et al.*, 2008) and plastid (for example, see Fuller *et al.*, 2006b; MacDonald *et al.*, 2007; Lep re *et al.*,

Correspondence: DJ Scanlan, School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK.

E-mail: d.j.scanlan@warwick.ac.uk

<sup>5</sup>Current address: School of Environmental Sciences, University of East Anglia, Norwich, UK

<sup>6</sup>Current address: Clermont Universit , Universit  Blaise Pascal, Laboratoire 'Microorganismes: G nome et Environnement', Clermont-Ferrand, France

<sup>7</sup>These authors contributed equally to this work.

Received 19 July 2012; revised 2 November 2012; accepted 8 November 2012; published online 31 January 2013

2009) small subunit rRNA genes have revealed a vast, previously unsuspected diversity within this assemblage. This focus is largely because of the increased recognition of PPEs as key contributors to phytoplankton biomass and primary production across various marine pelagic ecosystems (Li, 1994; Worden *et al.*, 2004; Richardson and Jackson, 2007; Cuvelier *et al.*, 2010; Jardillier *et al.*, 2010; Grob *et al.*, 2011). Moreover, PPEs have recently been shown to be key bacterivores controlling bacterioplankton abundance in both temperate (Zubkov and Tarran, 2008) and oligotrophic gyre (Hartmann *et al.*, 2012) waters. Furthermore, oceanic nitrogen:phosphorus (N:P) ratios are likely to be governed by plankton biogeography because of the varying nutrient metabolism of different groups (Weber and Deutsch, 2010). Thus, defining the abundance and understanding the diversity and global distribution of natural PPE populations is important for elucidating how specific PPE classes, with potentially different mixotrophy versus autotrophy capability, are linked to various physical and chemical parameters.

In the present study, abundances,  $\alpha$ -diversity (richness) and  $\beta$ -diversity (composition) of PPEs were assessed at a global scale (seven research cruises and four major ocean biomes) by different molecular methods, targeting the gene encoding the plastid 16S rRNA.

## Materials and methods

### Cruises

We analysed seven cruises, which took place between 2001 and 2007. The cruise tracks are depicted in Figure 1 with additional information (dates, geographic region and sampling points) summarised in Table 1. The cruise transects encompass (1) all four major ocean biomes (Longhurst, 2007): polar, westerly winds, trade winds and coastal boundary domains; (2) various temperature zones: arctic, temperate, subtropical and tropical; and (3) distinct oceanic regimes (as defined by

nutrient status): gyres, upwelling, equatorial and coastal regions.

### Sampling

Sampling procedures and nutrient analyses for the ARCTIC, VANC10MV and BEAGLE cruises (Figure 1 and Table 1) can be found in Not *et al.* (2005, 2008) and Bouman *et al.* (2006). In brief, samples analysed from the Arctic Ocean cruise were collected at six depths from the surface of the water column to ~60 m deep in August 2002. Indian Ocean samples were collected from the upper 800 m on the VANC10MV cruise during May–June 2003 between Cape Town (South Africa) and Port Hedland (Australia), passing through the south central Indian Ocean Gyre. The BEAGLE cruise circumnavigated the Southern Ocean between 20°S and 32.5°S, and sampled only at the surface of the water column. Because of the small amount of material collected on the BEAGLE cruise, cells collected from three alternate stations were combined for DNA extraction in order to provide enough material for analysis. Environmental samples were taken with a rosette equipped with Niskin bottles. For DNA extraction, 10 l of sea water was filtered first through a 47 mm diameter, 3  $\mu$ m pore size polycarbonate filter (Millipore, Billerica, MA, USA) and then onto a 47 mm diameter, 0.45  $\mu$ m pore size polysulfone filter (Supor450, Gelman Sciences, Ann Arbor, MI, USA) under gentle vacuum (10 mm Hg). The filters were transferred into 5 ml cryotubes containing 3 ml of DNA lysis buffer (0.75 M sucrose, 400 mM NaCl, 20 mM EDTA and 50 mM Tris, pH 9.0), flash-frozen in liquid nitrogen and stored at –80 °C. DNA was subsequently extracted from the filters as described previously (Fuller *et al.*, 2003). Sampling details for the other cruises included in the meta-analysis can be found in Fuller *et al.* (2006a), Lepère *et al.* (2009) and Kirkham *et al.* (2011a, b). We also include sequencing data from a time series taken from the Gulf of Naples, Mediterranean Sea (see McDonald *et al.*, 2007). Environmental parameters

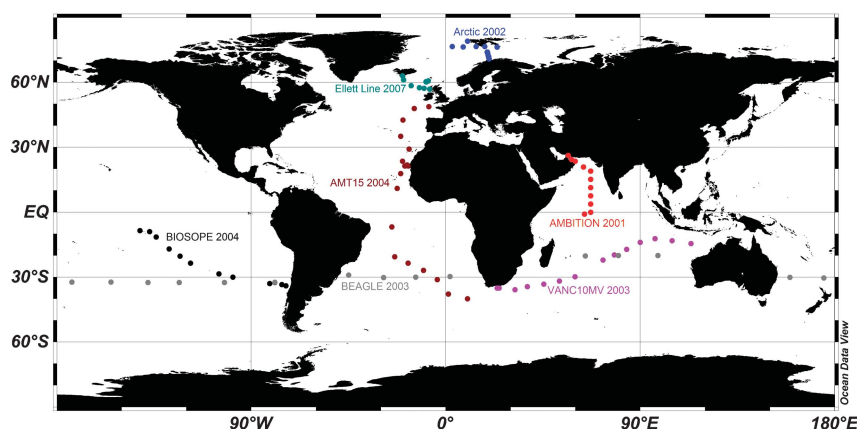


Figure 1 Schematic representation of the cruise tracks analysed in this study.

**Table 1** Details of cruises analysed in this study

<i>Cruise</i>	<i>Date</i>	<i>Geographic region</i>	<i>Latitude/ longitude</i>	<i>Sampling</i>	<i>Reference</i>
AMBITION	03/09/01 to 27/09/01	Arabian Sea	1°S–26°N, 56–67°E	11 stations, 6 depths	Fuller <i>et al.</i> (2006a)
AMT15	19/09/04 to 26/10/04	Atlantic Meridional Transect	48°N–40°S, 25°W–10°E	21 stations, 6 depths	Kirkham <i>et al.</i> (2011a)
ARCTIC	24/08/02 to 31/08/02	Arctic Ocean	70–79°N, 3–23°E	11 stations, 5 depths	Not <i>et al.</i> (2005) and this study
ELLETT	25/08/07 to 04/09/07	Northern North Atlantic	63–57°N, 20–7°W	9 Stations, 5–6 depths	Kirkham <i>et al.</i> (2011b)
BEAGLE	08/07/03 to 13/01/04	Circumnavigation of the Southern Hemisphere	20–32°S	15 Stations (surface only)	Bouman <i>et al.</i> (2006) and this study
BIOSOPE	27/10/04 to 06/12/04	Eastern South Pacific Ocean	8–34°S, 141–72°W	11 Stations 6 depths	Lepère <i>et al.</i> (2009)
VANC10MV	16/05/03 to 11/06/03	Indian Ocean	35–43°S, 24–113°E	15 stations, 6 depths	Not <i>et al.</i> (2008) and this study

measured for each cruise/station can be found in Supplementary Table 1.

#### *Flow cytometric analysis*

PPEs, *Prochlorococcus* and *Synechococcus* were enumerated by flow cytometry (FACSort, Becton Dickinson, Oxford, UK) using their characteristic pigment autofluorescence and size. The flow rate was calculated by adding a known concentration of 0.5 mm multi-fluorescent latex beads (Polysciences, Eppelheim, Germany) as an internal standard. Flow cytometry data were processed using CellQuest software (Becton Dickinson).

#### *PCR conditions*

PCR amplification of the 16S rRNA gene from environmental and control DNA samples for dot blot hybridisation and/or clone library construction were performed as described in Kirkham *et al.* (2011a) using the algal plastid biased primer PLA491F (Fuller *et al.*, 2006b) coupled with the general oxygenic phototroph primer OXY1313R (West *et al.*, 2001) to give an ~830 bp PCR product.

#### *Dot blot hybridisation analysis*

16S rDNA amplicons from environmental DNAs and control strains were purified, blotted onto nylon membranes and hybridised to algal class-specific oligonucleotide probes, following the method of Fuller *et al.* (2003). The oligonucleotide probes used for all cruises were: CHLA768, CHRY1037, CRYP862, EUST985, PAVL665, PELA1035, PING1024, PRAS826, PRYM666 and TREB708 targeting the plastids of Chlorarachniophyceae, Chrysophyceae, Cryptophyceae, Eustigmatophyceae, Pavlovophyceae, Pelagophyceae, Pinguicophyceae, Prasinophyceae clade VI (Prasinococcales), Prymnesiophyceae and Trebouxiophyceae, respectively (Fuller *et al.*, 2006a, b). Algal cultures used as controls were the

same as those used by Fuller *et al.* (2006a) and were obtained from the Roscoff Culture Collection (RCC, <http://www.sb-roscoff.fr/Phyto/RCC/>) and the Provasoli-Guillard National Center for Marine Algae and Microbiota (NCMA, formerly the CCMP, <https://ncma.bigelow.org/>). Final wash (or dissociation) temperatures (Td) for each probe were determined empirically (Fuller *et al.*, 2006b), following a previously described method (Fuller *et al.*, 2003). Hybridisation was quantified by using a Fujifilm FLA-5000 phosphorimager and Total Laboratory software (Phoretix, Newcastle, UK). The relative hybridisation of a given specific probe compared with that of the eubacterial probe to the control DNAs was averaged where more than one control DNA was used. Any sample giving a signal above 2% was considered above background.

#### *Construction of clone libraries*

PCR products were cloned into the TA vector pCR2.1-TOPO (Invitrogen, Paisley, UK) and screened by restriction fragment length polymorphism after digestion with *Hae*III and *Eco*RI as described previously (Kirkham *et al.*, 2011a).

#### *Data analysis*

Plastid 16S rRNA gene sequences obtained from the Arctic Ocean library have been deposited in Genbank under accession numbers GQ863885–GQ863899. Sequence accession numbers for the AMBITION (Arabian Sea), BIOSOPE (Pacific Ocean), AMT (Atlantic Ocean), Mediterranean Sea and Ellett Line (North Atlantic Ocean) can be found in Fuller *et al.* (2006a, b), McDonald *et al.* (2007), Lepère *et al.* (2009); Shi *et al.* (2011), Kirkham *et al.* (2011a) and Kirkham *et al.* (2011b), respectively.

The Margalef index (Hill *et al.*, 2003) was used for quantifying the richness ( $\alpha$ -diversity) of PPEs as, for some libraries, we lacked data for the abundance of clones associated with each operational taxonomic

unit, required for the use of other types of indices. Plastid 16S rRNA gene sequences obtained from cloning–sequencing were aligned and phylogenetic trees (neighbour-joining algorithm with Jukes–Cantor correction) were produced using ARB (Ludwig *et al.*, 2004). The resulting trees were used in Unifrac analysis (<http://bmf2.colorado.edu/unifrac/index.psp>; Lozupone and Knight, 2005) to compare  $\beta$ -diversity between libraries.

To explore relationships between environmental parameters and the distribution of PPEs measured by hybridisation of class-specific oligonucleotide probes, canonical correspondence analysis (CCA) was used (Ter Braak, 1986). Variables included chlorophyll *a*, concentrations of phosphate, nitrate and nitrite (measured together), salinity, depth, mixed layer depth, latitude, temperature and season. For each sample, two values were assigned for season. One (Spring) was based on the length of time from the autumn equinox and the second (Summer) on the length of time from the winter solstice. This allowed season to be treated as a continuous variable and made account of the hemisphere from which the sample was collected. The CCA (<http://cran.r-project.org/>) analysis identifies the linear combinations of class abundance variables explaining the most variation among the samples, constrained to be maximally related to weighted linear combinations of the environmental variables (Legendre and Legendre, 1998). Thus, the fitted analysis model identifies the strongest associations between class abundance and environmental variables. CCA plots were drawn using R software with biplot values of the environmental variables and eigenvalue weighted eigenvectors of dot blot hybridisation data. Associations between classes (from the dot blot hybridisation data) and each environmental variable can be identified by considering the orthogonal projections from each class mean point on the CCA plot onto the appropriate environmental variable vector (Legendre and Legendre, 1998). Pairwise two-sided Spearman correlation coefficients were calculated using R software using the Agricolae package to provide further support for the associations identified from the CCA plots.

## Results and discussion

### *Marine PPE abundance over large spatial scales*

PPEs were enumerated by flow cytometry across all transects (Figure 2 and Supplementary Table 1). The highest PPE abundances were found in the cool (12–15 °C), high chlorophyll *a* surface (0–10 m) waters in the northern Atlantic Ocean and Arctic Ocean, with a peak of  $3.9 \times 10^4$  cells per ml encountered in the Rockall trough of the extended Ellett Line transect (Figure 2 and Supplementary Table 1). The lowest abundances were found in the Arabian Sea transect (mean over the transect  $8.2 \times 10^2$  cells per ml), being below detection limits for much of the transect. PPE densities dropped from the surface to the DCM for all transects. Lowest densities were found in oceanic gyres (with an average of  $9.2 \times 10^2$  cells per ml for all of the gyre stations sampled) but with higher densities at the more nutrient-rich stations, as described previously (Worden and Not, 2008). *Prochlorococcus* was the dominant phototroph in oligotrophic regions of all the transects. Although PPEs are generally less abundant than picocyanobacteria, they represent the most abundant group in stations sampled north of 74°N in the Arctic cruise as well as in stations in the easternmost part of the Indian Ocean cruise. These findings support the belief that PPEs play a significant role in the primary production of polar ecosystems, replacing their cyanobacterial counterparts, which numerically dominate at lower latitudes (see, for example, Zubkov *et al.*, 2003). Although there is a general trend in the PPE/total picophytoplankton (that is, PPEs + picocyanobacteria) ratio increasing systematically with increasing latitude and decreasing temperature (Bouman *et al.*, 2012), a considerable amount of variability is observed across the range of latitudes and temperatures. It is important to remember that even at very low cell abundances, PPEs are now known to contribute significantly to marine primary production because of a multi-factorial effect of greater biovolume, higher growth rates and high grazing mortality rates (Li, 1994; Worden *et al.*, 2004; Jardillier *et al.*, 2010; Grob *et al.*, 2011).

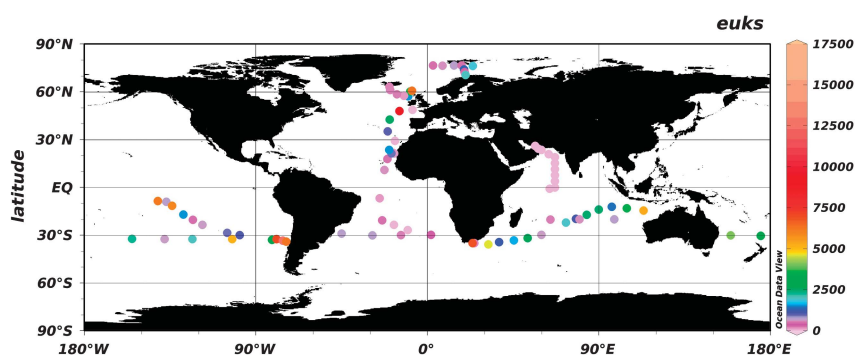


Figure 2 Global distribution patterns of PPE abundances, at all depths sampled, determined by flow cytometry.



*PPE  $\alpha$ -diversity*

Clone sequence data of the 16S rDNA from all the transects studied here (except the BEAGLE and Indian Ocean transects) representing 31 clone libraries allowed us to calculate species richness values ( $\alpha$ -diversity) using the Margalef index (Dmg) (Table 2). A relatively high variation in this index was seen between sites (Table 2) with, on average, samples taken from the Gulf of Naples, Mediterranean Sea, showing the highest diversity (Dmg = 7.3). This high richness was positively correlated with temperature ( $r=0.84$ ,  $P<0.05$ ) and low nutrient concentrations ( $r=0.71$ ,  $P<0.05$ ), that is, oligotrophic stations over the time series. Conversely, the lowest richness was observed along the Arabian Sea and Atlantic Ocean transects (average Dmg = 0.97 and 1.97, respectively) and seems to coincide with low PPE abundances. On the BIOSOPE transect (the only cruise where different depths were analysed for  $\alpha$ -diversity), PPE species richness decreased with depth. Similar conclusions were made by Schnetzer *et al.* (2011), but for total microbial eukaryote diversity (again using environmental gene libraries), with depth negatively influencing species richness in the eastern North Pacific.

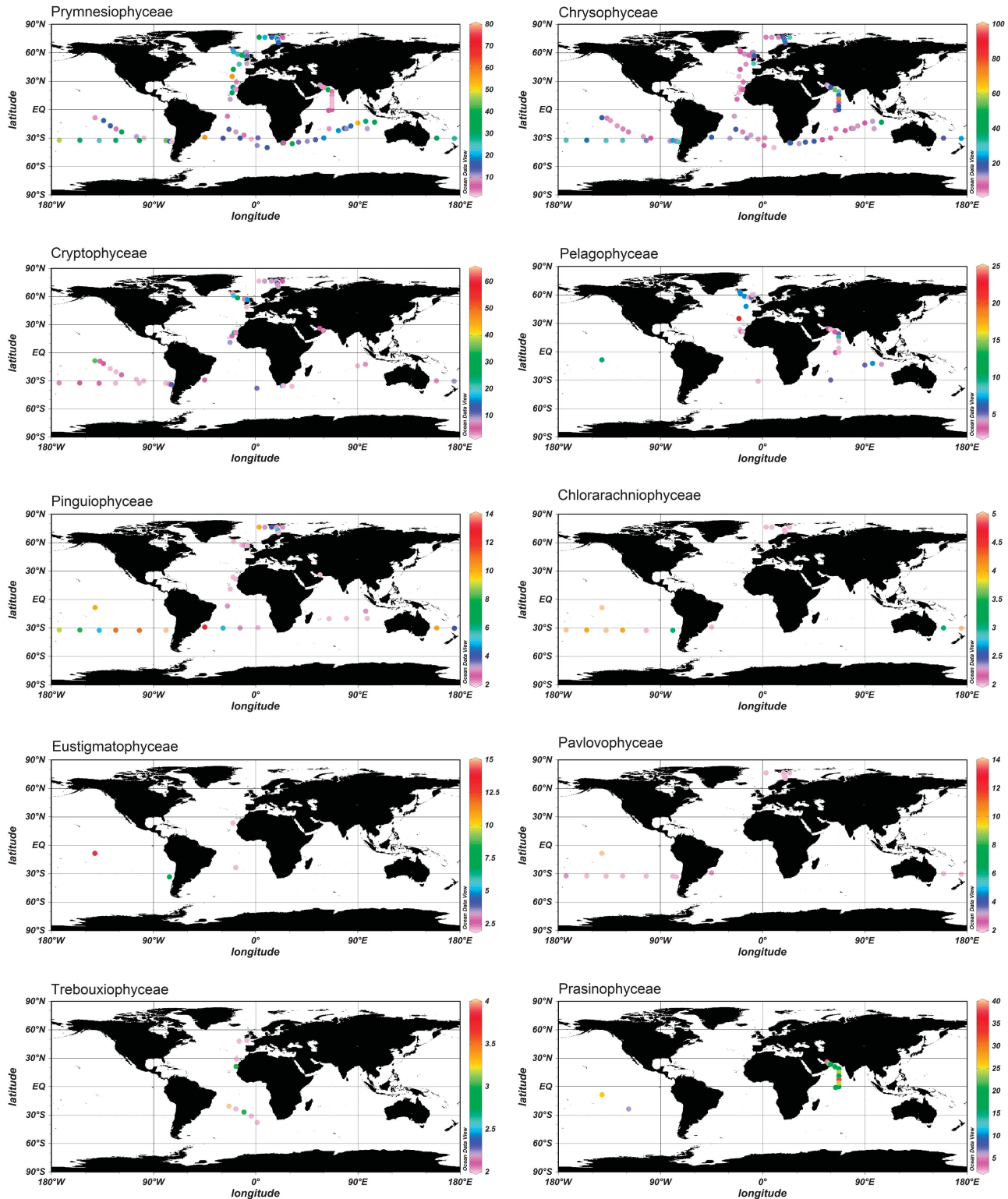
*PPE  $\beta$ -diversity*

**Global PPE class distributions.** Plastid 16S rRNA oligonucleotide probes used for dot blot hybridisation analysis revealed specific global distribution patterns for each PPE class detected (Figure 3). The classes Prymnesiophyceae and Chrysophyceae were globally important across the range of ocean environments analysed (Figures 3 and 4). These classes were detected in every sample analysed and on average comprise 78% of the total relative hybridisation values obtained with the 10 PPE class-specific probes used over all 7 transects. Interestingly, these two classes have complementary distribution patterns across several transects (AMT, BIOSOPE, ARCTIC and AMBITION). The high Prymnesiophyceae signal detected across all ocean basins (Supplementary Table 3) supports the observation that 19'-hexanoyloxyfucoxanthin, a prymnesiophyte-specific pigment (though also present in a few other Heterokont algae, see Andersen, 2004), often dominates oceanic pigment analyses (Not *et al.*, 2008; Liu *et al.*, 2009). Recent fluorescent *in situ* hybridisation studies confirm the high abundance of these pico-prymnesiophytes in the

**Table 2** Photosynthetic picoeukaryote (PPE)  $\alpha$ -diversity calculated using the Margalef index (DMg)

Cruises	Clone libraries	Depth (m)	Latitude	Longitude	Margalef index (DMg)	DMg mean
AMT 15	1B ●	10	48.74	-7.85	3.57	1.97
	15B ●	10	17.83	-20.89	0.82	
	27B ●	30	-23.56	-17.49	0.27	
	33B ■	5	-37.83	1.23	3.21	
ELLETT LINE	IB4 ●	32	58.5	-16	3.57	4.00
	EG3 ●	10	60.25	-9.01	4.43	
BIOSOPE	MAR1 ■	15	-8.38	-141.23	5.76	2.94
	STB1 ●	25	-11.74	-134.10	3.05	
	STB6 ●	55	-20.44	-122.91	3.30	
	STB6 ●	180	-20.44	-122.91	1.48	
	STB7 ●	40	-22.04	-120.35	4.07	
	STB7 ●	175	-22.04	-120.35	2.40	
	STB11 ●	0	-27.76	-107.27	3.00	
	STB11 ●	5	-27.76	-107.27	4.42	
	STB11 ●	200	-27.76	-107.27	2.80	
	STB14 ●	5	-30.03	-98.39	4.50	
	STB14 ●	150	-30.03	-98.39	2.35	
	STB17 ●	20	-32.39	-86.78	4.04	
	STB17 ●	30	-32.39	-86.78	1.64	
	STB17 ■	70	-32.39	-86.78	1.39	
	UPW1 ★	5	-33.96	-73.37	4.20	
	UPW1 ★	35	-33.96	-73.37	3.00	
Mediterranean Sea	MC597 ●	surface	40.81	14.31	6.35	7.31
	MC601 ●	surface	40.81	14.31	2.34	
	MAMA64 ●	surface	40.81	14.31	4.67	
	250304 ●	surface	40.81	14.31	11.83	
	MC615 ●	surface	40.81	14.31	9.69	
	MC622 ●	surface	40.81	14.31	8.98	
Ambition	AS2 ■	50	-0.001	66.99	1.30	0.97
	AS10 ★	50	24.33	58.16	0.65	
Arctic	A 50 ■	60	76.5	3	2.30	2.3

Coloured symbols illustrate the temperature and trophic status of the samples from which clone libraries were constructed. Circle, oligotrophic (Phosphate concentration  $<10 \text{ mg m}^{-3}$ ); square, mesotrophic (Phosphate concentration  $10\text{--}20 \text{ mg m}^{-3}$ ); star, eutrophic (Phosphate concentration  $>20 \text{ mg m}^{-3}$ ); blue, very low temperature ( $<10^\circ\text{C}$ ); green, low temperature ( $10\text{--}14.99^\circ\text{C}$ ); yellow, medium temperature ( $15\text{--}19.99^\circ\text{C}$ ); orange, high temperature ( $20\text{--}24.99^\circ\text{C}$ ); red, very high temperature ( $\geq 25^\circ\text{C}$ ).



**Figure 3** Global distribution patterns of specific PPE classes, at all depths sampled, as determined by dot blot hybridisation analysis.

Atlantic Ocean (Jardillier *et al.*, 2010; Grob *et al.*, 2011; Kirkham *et al.*, 2011b), Indian Ocean (Not *et al.*, 2008) and in open ocean regions of the Arctic Ocean (Not *et al.*, 2005). However, Prymnesiophyceae have been

consistently underestimated in previous amplification-based studies using primers targeting the nuclear 18S rRNA gene, potentially a result of PCR bias (Liu *et al.*, 2009). In contrast, genetic surveys targeting the plastid

16S rRNA gene have shown a high diversity of this group in both open ocean (Lepère *et al.*, 2009; Kirkham *et al.*, 2011a, b) and coastal waters (McDonald *et al.*, 2007). Together, these data suggest large global distributions of pico-sized prymnesiophytes, which contribute significantly to marine primary production (Cuvelier *et al.*, 2010; Jardillier *et al.*, 2010) even though they mostly lack cultured representatives. The success of these organisms could be because of mixotrophic behaviour (that is, an ability to supplement their phototrophic physiology by preying on bacterioplankton). Indeed, recent evidence from natural environments suggests that PPEs can contribute greatly to bacterivory (Zubkov and Tarran, 2008; Sanders and Gast, 2011; Hartmann *et al.*, 2012). The nutritional flexibility offered by mixotrophy gives a significant competitive advantage over both purely phototrophic and aplastidic cells under different light (depth) and nutrient regimes.

Relative hybridisation values for Chrysophyceae (average 15.1% relative hybridisation across all cruises) were almost as high as those of Prymnesiophyceae (Supplementary Table 3) with peak values of 39% relative hybridisation at the westerly most station (Z59) sampled in the Arctic Ocean and 49% relative hybridisation at the eastern end of the Indian Ocean transect. Interestingly, Chrysophyceae signals were detectable to the bottom of the water column sampled (between 0 and 800 m according to the cruise), whereas most other PPEs classes were not detected below 150 m. The exception was the Trebouxiophyceae class that at some stations in the Indian Ocean was found down to 200 m (outside the gyre) and 800 m (inside the gyre). The presence of Chrysophyceae and Trebouxiophyceae in these deep waters well below the photic zone might be explained by their mixotrophic growth potential (the advantages of which were explained above).

Cryptophyceae were frequently observed in surface coastal waters of the Pacific and Atlantic Oceans. However, relative hybridisation values rarely exceeded 10% (Figure 3). Their low abundance in open ocean waters has also been observed using fluorescent *in situ* hybridisation analysis (for example, 1–3% of phototrophic cells in the Indian Ocean (Not *et al.*, 2008) and <1% in the northern North Atlantic (Kirkham *et al.*, 2011b)) and by 18S and 16S rRNA gene sequencing (Not *et al.*, 2008; Lepère *et al.*, 2009; Shi *et al.*, 2009; Kirkham *et al.*, 2011a). This class appears to be mostly restricted to coastal waters (see, for example, Romari and Vaultot, 2004). Moreover, cultured cryptophytes are generally larger than 5 µm, although some groups have been documented within the small fraction (Vaultot *et al.*, 2008). Members of the Pinguiphyceae were fairly widespread, reaching relative hybridisation values of ≥2% in at least one sample of every cruise analysed except the Arabian Sea, and reaching up to 8% relative hybridisation in mesotrophic waters of the Pacific Ocean and the four easternmost stations of the Indian Ocean transect. They were also fairly

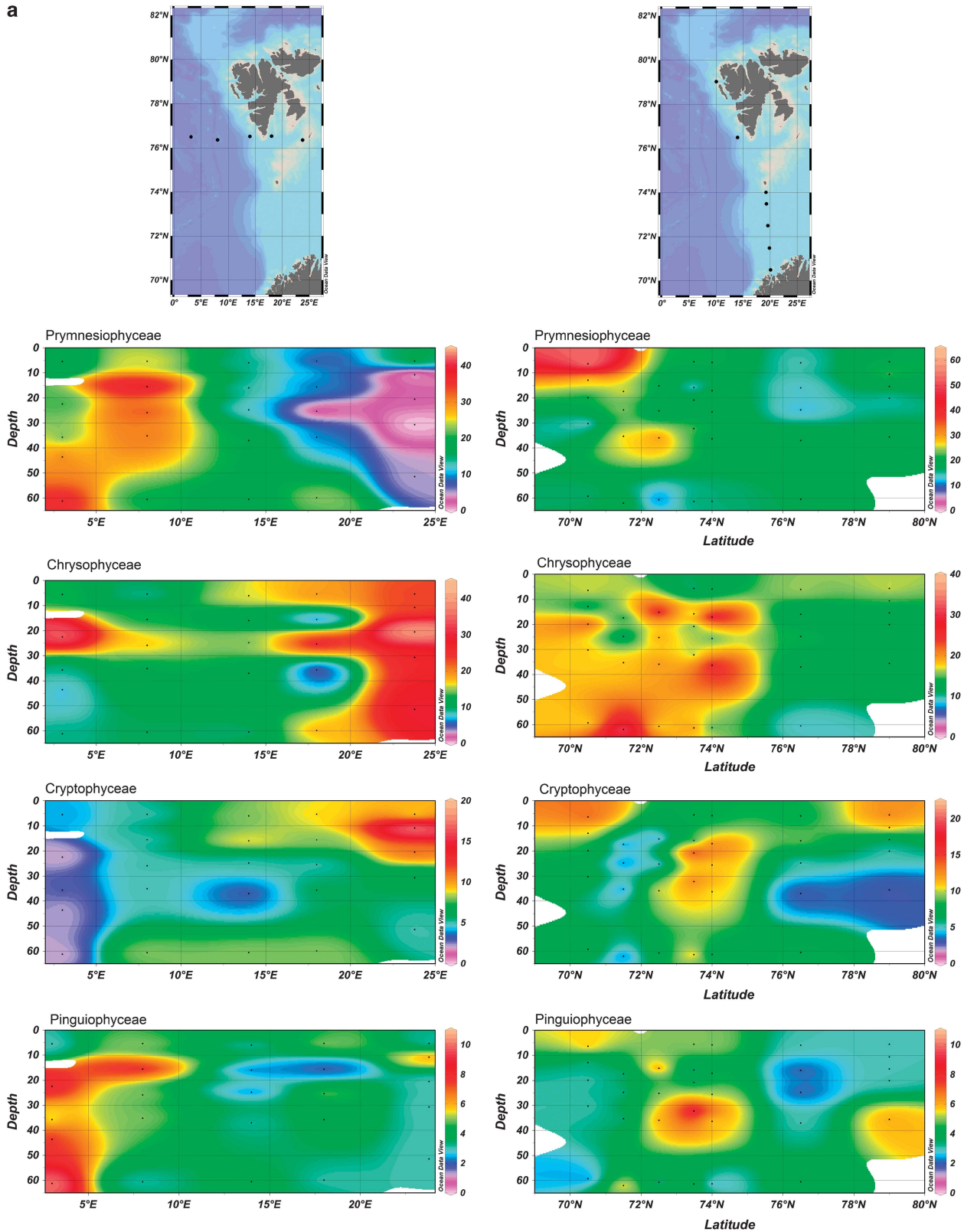
well represented in the western part of the Arctic Ocean transect (Figure 4a).

Pelagophyceae, Pavlovophyceae, Eustigmatophyceae, Chlorarachniophyceae, Trebouxiophyceae and Prasinophyceae clade VI were only detected sporadically (Figure 3). Curiously, despite being readily cultured (Le Gall *et al.*, 2008), Pelagophyceae were not detected in over 80% of the samples analysed, including the entire Arctic Ocean and BEAGLE cruises. This may be because of PCR bias given that Shi *et al.* (2011) showed the presence of many sequences related to Pelagophyceae using a different plastid 16S rRNA gene primer set on samples from the Pacific Ocean. Moreover, Not *et al.* (2008) found that pelagophytes contributed up to 40% of photosynthetic pigments in the Indian Ocean, whereas Jardillier *et al.* (2010) reported that PPEs of size  $1.8 \pm 0.1 \mu\text{m}$  comprised 14–57% Pelagophyceae via fluorescent *in situ* hybridisation analysis across a range of samples in the subtropical North Atlantic.

*PPE class distributions: relationships to environmental variables.* CCA was performed on each cruise separately, revealing associations between PPE class distributions and physicochemical environmental parameters (see Supplementary Figure 1) as well as previously published cruise transects (Lepère *et al.*, 2009; Kirkham *et al.*, 2011a, b). In these separate CCA analyses, the percent change in composition explained by the measured environmental variables was on average 50%. CCAs were also performed for the global data set ( $n=239$ ), excluding the Indian Ocean cruise and the Gulf of Naples time series for which few chemical parameter measurements were available (Figure 5). Data for 11 environmental parameters were obtained for all the other cruises (Supplementary Table 1), comprising season (spring=length of time from autumn equinox; and summer=length of time from winter solstice), chlorophyll *a*, phosphate ( $\text{PO}_4$ ), nitrate + nitrite concentration ( $\text{NO}_3 + \text{NO}_2$ ), nitrate + nitrite:phosphate ratio, salinity, depth, mixed layer depth, latitude and temperature. Only 20.6% of the total variation in dot blot hybridisation values could be explained by the measured environmental parameters. This is likely to be due, in part, to the very large degree of variation observed over such a large-scale study, and the comparatively small number of variables for which data had been collected for all the cruises. Furthermore, the lack of resolution beyond the class level is also likely to have contributed to the large proportion of variation in dot blot hybridisation data that was unexplained by the CCA.

Interestingly, the observed complementary distribution of Prymnesiophyceae and Chrysophyceae (Figure 4; Kirkham *et al.*, 2011a) is confirmed by the CCA plot where the two classes are mirrored (Figure 5). Based on the orthogonal projections from the class coordinates on the first two canonical axes from the overall CCA (Figure 5), higher abundance of Chrysophyceae is associated with samples





**Figure 4** Dot blot hybridisation data showing the distribution of specific PPE classes along (a) the Arctic Ocean transect, plotted by longitude (left) and latitude (right), (b) along the Indian Ocean transect and (c) along the BEAGLE transect, encompassing samples taken in surface waters of the Pacific Ocean, Atlantic Ocean and Indian Ocean. The three dots in red circle correspond to three CTDs from where DNA was pooled. Contour plots (a, b) indicate the percent relative hybridisation (as a proportion of all products amplified by primers PLA491F and OXY1313R). The y-axes plot the depth (m) down each water column, and the x-axes plot the distance along the cruise by longitude (left) and latitude (right). Black dots represent sampling points.



b

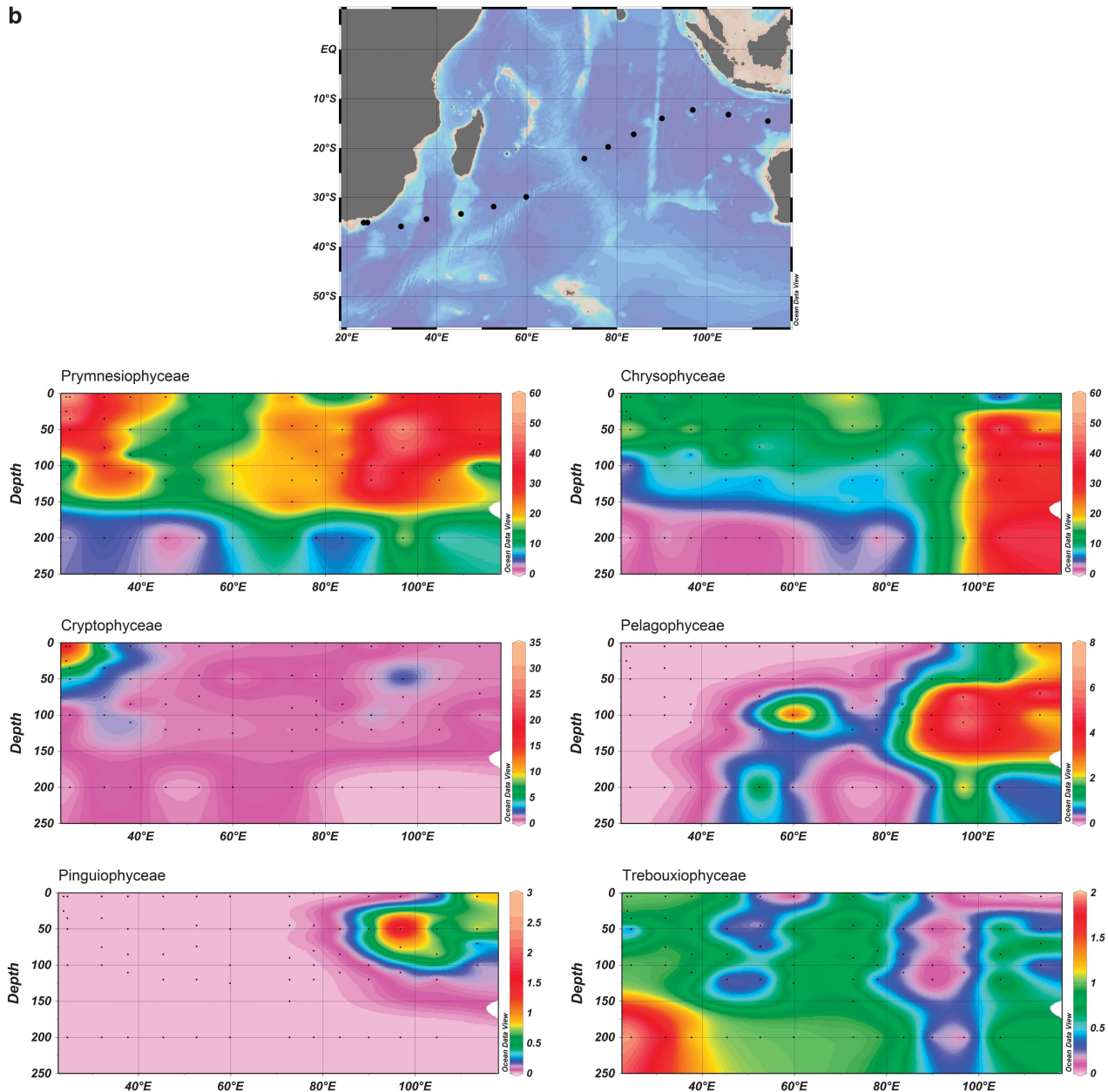


Figure 4 Continued.

collected in early spring (more positive values on the 'spring' vector), lower latitudes (more negative values on the latitude vector), higher temperatures, lower N:P ratios and higher  $\text{PO}_4$  concentrations. Conversely, higher abundance of Prymnesiophyceae is associated with samples taken later in the year, at higher latitudes, higher  $\text{NO}_3 + \text{NO}_2$  concentrations and higher N:P ratios. Based on the orthogonal projections that can be inferred from Figure 5, Pelagophyceae, Prymnesiophyceae and Cryptophyceae are associated with samples collected in the late autumn, with low concentrations of  $\text{PO}_4$  and higher N:P ratios. In contrast, Prasinophyceae, Trebouxiophyceae, Eustigmatophyceae and Chrysophyceae are associated with samples collected in the

early spring, with higher concentrations of  $\text{PO}_4$  and lower N:P ratios. Some of these associations are supported by considering pairwise Spearman correlations, for example, Chrysophyceae had a correlation coefficient of 0.35 with spring ( $P < 0.01$ ),  $-0.28$  with latitude ( $P < 0.01$ ), 0.36 with temperature ( $P < 0.01$ ) and  $-0.24$  with N:P ratio ( $P < 0.01$ ). Prymnesiophyceae had a correlation coefficient of  $-0.3$  with spring ( $P < 0.01$ ), 0.21 with latitude ( $P < 0.01$ ) and 0.2 with N:P ratio ( $P < 0.01$ ). However, it should be noted that these simple correlation calculations ignore the true complexity of the associations among the class distributions and environmental variables that is captured by the overall CCA model.

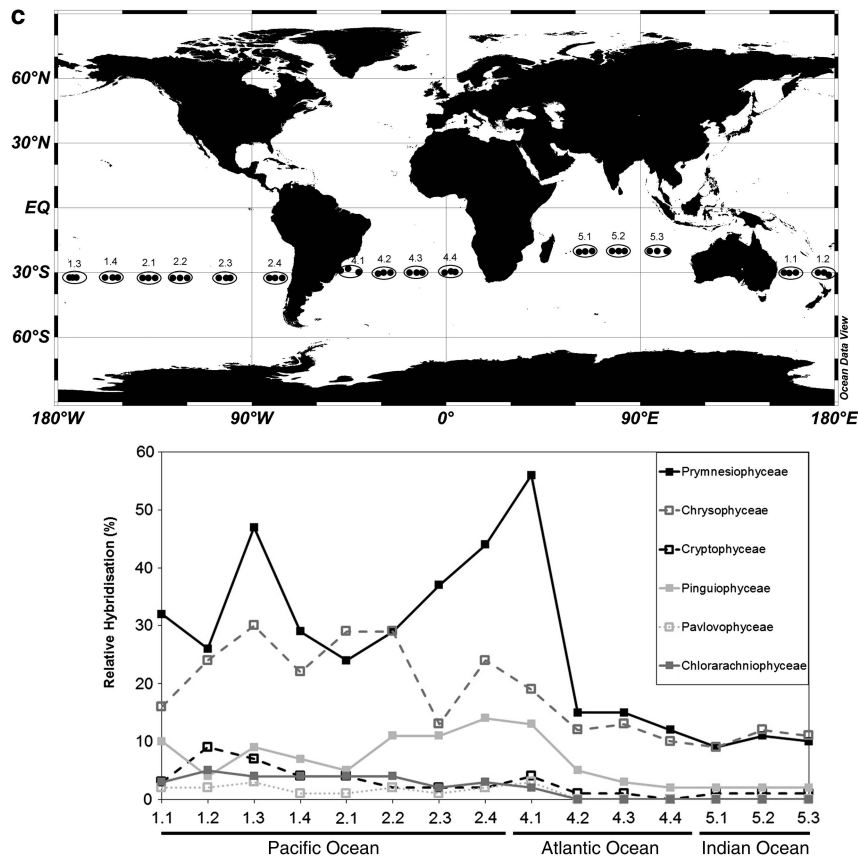


Figure 4 Continued.

Considering the variables of phosphate and  $\text{NO}_3 + \text{NO}_2$  concentrations in more detail, peaks in relative hybridisation values for the Prymnesiophyceae tended to be linked with lower  $\text{PO}_4$  concentrations. Large prymnesiophytes in the Ross Sea are responsible for the export of higher N:P organic matter than that attributed to diatoms (Arrigo *et al.*, 1999), and it is possible that pico-prymnesiophytes might co-vary with these larger prymnesiophytes. Samples with peak Chrysophyceae relative hybridisation values tended to have higher  $\text{PO}_4$  values (especially along the AMBITION, BIOSOPE and AMT transects). Marine production is constrained by nutrient availability, and phytoplankton N:P ratios differ with group. For example, green algae have higher ratios than red algae and growth strategies also alter this ratio (Falkowski *et al.*, 2004; Arrigo, 2005). Generalist strategies have a near Redfield N:P ratio of 16:1, opportunistic strategies rely on low N:P ratios and survivalist strategies rely on high N:P ratios (Arrigo, 2005). The average N:P ratio of the water samples analysed in this study was calculated to be 14:1 by linear regression, similar to the study of Tyrell (1999) that reported a 15:1 N:P ratio. When considering only samples for which Prymnesiophyceae relative hybridisations were >30%, the average N:P ratio was 25:1, whereas

when considering only samples for which Chrysophyceae relative hybridisation values were >30%, the average N:P ratio was 12:1. According to Weber and Deutsch (2010), phytoplankton may influence the N:P ratio of their surrounding water because of their differing metabolism. Alternatively, phytoplankton may be adapted to the N:P ratio of their surroundings, resulting in niche differentiation that may underlie the distribution patterns of the Prymnesiophyceae and the Chrysophyceae (see Litchman and Klausmeier, 2008). However, it should be considered that both classes had peaks at very low concentrations of these nutrients and it is likely that other factors in complex interactions also exerted influence over their distributions.

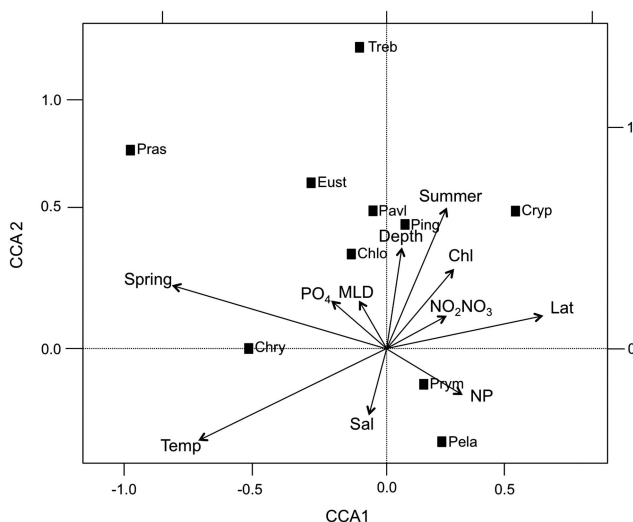
*Plastid 16S rRNA gene sequences: global distribution analyses.* In order to better understand global distributions of PPEs at a higher taxonomic level, all currently available plastid 16S rRNA gene sequences (this study; Fuller *et al.*, 2003; McDonald *et al.*, 2007; Lepère *et al.*, 2009; Kirkham *et al.*, 2011a, b; Shi *et al.*, 2011) obtained using the PLA491F/OXY1313R primer pair were analysed via Unifrac. This primer set is considered to best encompass the extent of PPE diversity currently observed using other 16S or 18S rRNA

primer sets (for example, see Moon van der Staay *et al.*, 2001; Shi *et al.*, 2011) despite some known biases (see McDonald *et al.*, 2007). It should be borne in mind that the number of plastid genomes per cell may also bias the data (see Maguire *et al.*, 1995). Moreover, we would suggest that a cloning–sequencing approach allows targeting of mainly abundant taxa, leaving the rare taxa unseen (Pedros-Alio, 2006). As such, only the distribution of abundant taxa is likely to be deciphered here, and hence we cannot exclude that the distribution of rare taxa within the same group could be different.

Unifrac analysis was performed on the total PPE genetic diversity as well as within specific PPE classes: Prymnesiophyceae, Chrysophyceae, Prasinophyceae and Cryptophyceae (Figure 6 and Supplementary Table 2). The *P*-value matrices using the total PPE data set that compared each library with each other showed a significant difference in the phylogenetic composition of PPEs between most of the different ocean basins (Supplementary

Table 2a). However, some clustering of libraries from environments with similar nutrient status was noted despite the distance between locations (Figure 6a). For example, libraries from oligotrophic stations in the Atlantic Ocean, Pacific Ocean and Gulf of Naples (Mediterranean Sea) (AMT\_27, Biosope\_STB11\_0, Biosope\_STB6\_55 and MC622) clustered closely together concomitant with low  $\text{NO}_3 + \text{NO}_2$  and  $\text{PO}_4$  concentrations and very similar water column temperatures ( $>20^\circ\text{C}$ ). The Arctic Ocean and south Atlantic Ocean samples also branched very closely and share similar salinities (34.93 and 34.8 PSU), low temperatures ( $<13^\circ\text{C}$ ), richness values indicating moderate species richness (compared with the range of values of the other libraries analysed) and similar Chrysophyceae abundances according to dot blot hybridisation data. Conversely, it is clear that the genetic similarity of different libraries is not only based on temperature and trophic status as libraries of similar status are well spread on the tree; for example, the Pacific (BIOSOPE) and Atlantic (AMT) gyre samples are genetically very different from each other, as are sequences from clone libraries derived from upwelling regions. In addition, libraries from dissimilar regions of the Atlantic (AMT1B), Arabian Sea (AS2) and Pacific Ocean upwelling region (UPW1\_35m) clustered closely together despite having very different temperature, mixing regimes and nutrient status. However, noteworthy is that all three sites have comparatively high chlorophyll *a* concentrations (ranging between 0.28 and  $1.37\text{ mg m}^{-3}$ ) and moderate species richness determined by the Margalef index. Even so, resolving the combination of parameters responsible for structuring the PPE community is likely to be much more complex than for picocyanobacterial communities (Zwirgmaier *et al.*, 2008) given the differences of taxonomic resolution between groups (picocyanobacteria have been studied at the genera level).

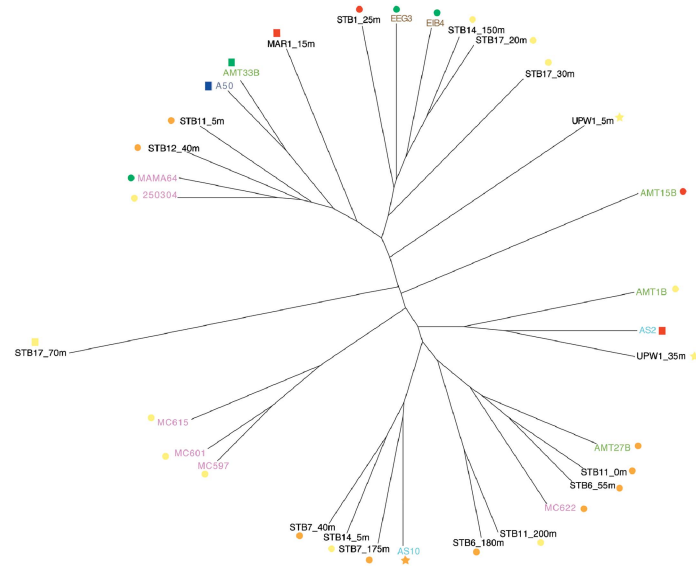
For some libraries whose overall similarity was close, substantial genetic variation within a class was observed. For example, two clone libraries from the Ellett Line cruise (EEG3 and EIB4), which are geographically adjacent stations of comparable trophic and temperature status, present a very similar overall composition (Figure 6a and Supplementary Table 2). However, their Prymnesiophyceae sequences are genetically distinct (Supplementary Table 2b). Specific lineages observed within the Prymnesiophyceae (see, for example, Lepère *et al.*, 2009) likely include various ecotypes subject to different long- and short-term ecological constraints (Worden and Not, 2008). In



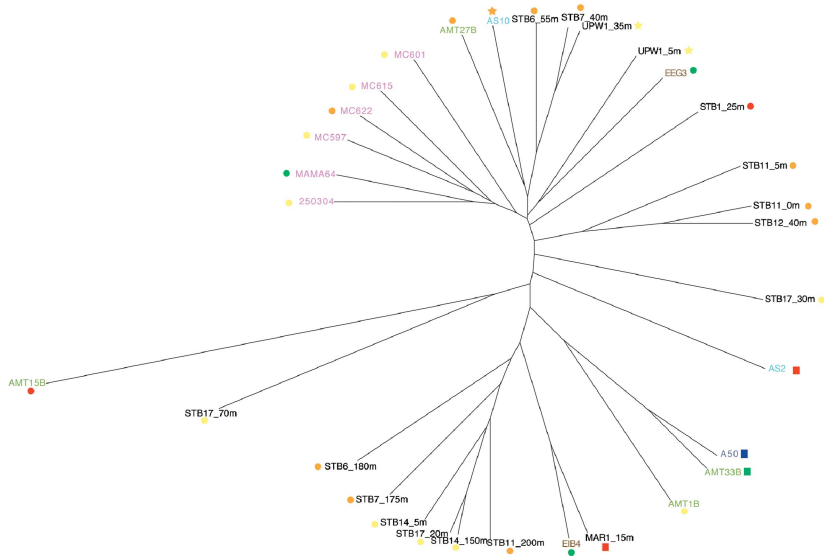
**Figure 5** Canonical correspondence analysis plot using relative hybridisation values (%) detected for all cruises except the Indian Ocean transect. PPE classes are as follows: Prymnesiophyceae (Prym), Chrysophyceae (Chry), Cryptophyceae (Cryp), Pinguicophyceae (Ping), Pelagophyceae (Pela), Eustigmatophyceae (Eust) and Trebouxiophyceae (Treb). Variables are season: length of time from autumn equinox (Spring), season: length of time from winter solstice (Summer), chlorophyll *a* (Chl), phosphate ( $\text{PO}_4$ ), nitrate + nitrite concentration ( $\text{NO}_2\text{NO}_3$ ), nitrate + nitrite:phosphate ratio (NP), salinity (Sal), depth (Depth), mixed layer depth (MLD), latitude (Lat) and temperature (Temp). The x-axis explains 9.3% and the y-axis explains 3.8% of the variation in dot blot hybridisation data.

**Figure 6** Unifrac analysis illustrating the genetic similarity of clone libraries based on sequences related to (a) all PPE classes, (b) Prymnesiophyceae and (c) Chrysophyceae. Coloured symbols illustrate the temperature and trophic status of the samples from which clone libraries were constructed. Circle, oligotrophic (Phosphate concentration  $<10\text{ mg m}^{-3}$ ); square, mesotrophic (Phosphate concentration  $10\text{--}20\text{ mg m}^{-3}$ ); star, eutrophic (Phosphate concentration  $>20\text{ mg m}^{-3}$ ); blue, very low temperature ( $<10^\circ\text{C}$ ); green, low temperature ( $10\text{--}14.99^\circ\text{C}$ ); yellow, medium temperature ( $15\text{--}19.99^\circ\text{C}$ ); orange, high temperature ( $20\text{--}24.99^\circ\text{C}$ ); red, very high temperature ( $\geq 25^\circ\text{C}$ ).

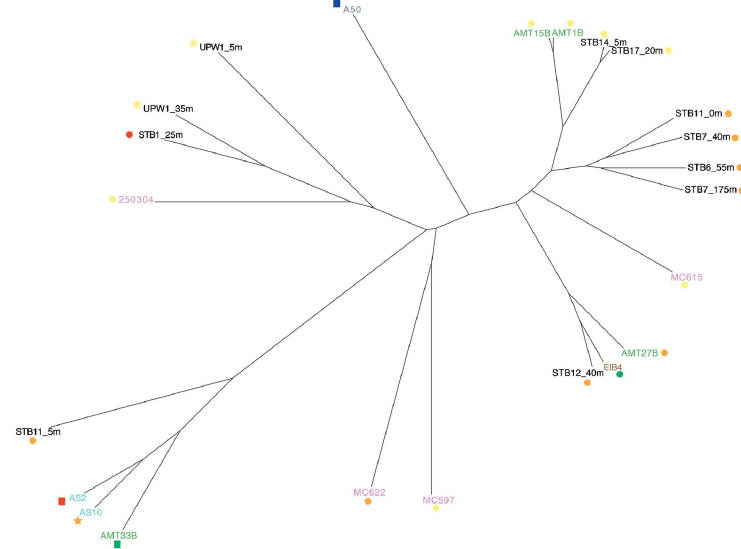
**a**



**b**



**c**





this respect, metagenomic analysis of a natural picoplanktonic population revealed a mosaic gene repertoire including specific adaptations for growth in oligotrophic environments (Cuvelier *et al.*, 2010), conditions that may represent a driver of niche differentiation. Furthermore, light availability has been shown to be associated with niche partitioning in prokaryotic picophytoplankton (Moore *et al.*, 1998; West and Scanlan, 1999) and within the prasinophyte genus *Ostreococcus* (Viprey *et al.*, 2008), and may also drive the niche adaptation process in the Prymnesiophyceae. Interestingly, the *P*-value matrix for Chrysophyceae and Cryptophyceae showed almost no significant difference in their phylogenetic composition between the samples analysed here (Supplementary Tables 2c and e). This may suggest that these classes contain much less variation than the Prymnesiophyceae and Prasinophyceae (Supplementary Tables 2b and d) and may instead contain widespread lineages (Figures 6b and c) with generalist growth strategies rather than differently niche-adapted lineages.

## Conclusions

This study further demonstrates the importance of pico-sized representatives of the Prymnesiophyceae and Chrysophyceae across all major ocean basins. At the global scale, our data showed an inverse distribution pattern of these two classes potentially associated with differing N:P requirements. On a vertical scale, our data also showed that although the  $\alpha$ -diversity of PPEs decreased with depth, their community structure ( $\beta$ -diversity) did not significantly change over this factor.

According to the cosmopolitan view of the microbial world, community structure (richness, diversity and composition) is likely similar in habitats that are alike, whereas differentiated microbial communities exist along an environmental gradient (Green and Bohannan, 2006). Although the sampling was not exhaustive, and many more taxa (especially low frequency and rare ones) might be present at the sampling sites, Unifrac analysis revealed at higher resolution than the class level that differences in  $\beta$ -diversity were greater between basins (interbasins) than within a basin (intra-basin). However, sampling in the Mediterranean and South Pacific regions showed significantly different PPE  $\beta$ -diversity along their time series or transect, respectively. These inter- and intra-basin changes in overall PPE  $\beta$ -diversity were linked to Prymnesiophyceae and Prasinophyceae composition, whereas Chrysophyceae seemed to be phylogenetically quite similar in all libraries. Within classes, lineages adapted to specific conditions, for example, those encompassing high and low light-adapted ecotypes (Rodríguez *et al.*, 2005) likely occur (for example, the prasinophyte *Ostreococcus*; Demir-

Hilton *et al.*, 2011). Further study of PPE distribution patterns at a higher taxonomic resolution is thus required and renewed efforts are needed to obtain into culture many PPE representatives only known from environmental gene sequencing.

## Acknowledgements

We thank the Captains and crews of all the cruises analysed. This study was supported by the UK Natural Environment Research Council via a PhD studentship award to ARK and NERC Grant NE/C003160/1 to DJS. Sequencing was performed at the NERC Biomolecular Analysis Facility in Edinburgh.

## References

- Andersen RA. (2004). Biology and systematics of heterokont and haptophyte algae. *Am J Bot* **10**: 1508–1522.
- Arrigo KR, Robinson DH, Worthen DL, Dunbar RB, DiTullio GR, VanWoert M *et al.* (1999). Phytoplankton community structure and the drawdown of nutrients and CO<sub>2</sub> in the Southern Ocean. *Science* **283**: 365–367.
- Arrigo KR. (2005). Marine microorganisms and global nutrient cycles. *Nature* **437**: 349–355.
- Bouman HA, Lepère C, Scanlan DJ, Ulloa O. (2012). Phytoplankton community structure in a high-nutrient, low-chlorophyll region of the eastern Pacific Subantarctic region during winter-mixed and summer-stratified conditions. *Deep Sea Res* **1** **69**: 1–11.
- Bouman HA, Ulloa O, Scanlan DJ, Zwirgmaier K, Li WK, Platt T *et al.* (2006). Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**: 918–921.
- Caron DA. (2009). Past President's address: protistan biogeography: why all the fuss? *J Eukaryot Microbiol* **56**: 105–112.
- Cuvelier ML, Allen A, Monier A, McCrow JP, Messié M, Tringe SG *et al.* (2010). Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *Proc Natl Acad Sci USA* **33**: 14679–14684.
- Demir-Hilton E, Sudek S, Cuvelier ML, Gentemann CL, Zehr JP, Worden AZ. (2011). Global distribution patterns of distinct clades of the photosynthetic picoeukaryote *Ostreococcus*. *ISME J* **5**: 1095–1107.
- Falkowski PG, Schofield O, Katz ME, Van de Schootbrugge B, Knoll AH. (2004). Why is the land green and the ocean red? In: Thierstein H, Young JR (eds) *Coccolithophores—From Molecular Processes to Global Impact*. Elsevier: Amsterdam, pp 429–453.
- Finlay BJ, Esteban GF, Brown S, Fenchel T, Hoef-Emden K. (2006). Multiple cosmopolitan ecotypes within a microbial eukaryote morphospecies. *Protist* **157**: 377–390.
- Finlay BJ, Fenchel T. (2004). Cosmopolitan metapopulations of free-living microbial eukaryotes. *Protist* **155**: 237–244.
- Foissner W. (2006). Biogeography and dispersal of microorganisms: a review emphasizing protists. *Acta Protozool* **45**: 111–136.
- Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL *et al.* (2008). A latitudinal diversity

- gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105**: 7774–7778.
- Fuller NJ, Campbell C, Allen DJ, Pitt FD, Zwirgmaier K, Le Gall F *et al.* (2006b). Analysis of photosynthetic picoeukaryote diversity at open ocean sites in the Arabian Sea using a PCR biased towards marine algal plastids. *Aquat Microb Ecol* **43**: 79–93.
- 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, Tarran GA, Cummings DG, Woodward EMS, Orcutt KM, Yallop M *et al.* (2006a). Molecular analysis of photosynthetic picoeukaryote community structure along an Arabian Sea transect. *Limnol Oceanogr* **51**: 2502–2514.
- Green J, Bohannan BJM. (2006). Spatial scaling of microbial biodiversity. *Trends Ecol Evol* **21**: 501–507.
- Grob C, Hartmann M, Zubkov MV, Scanlan DJ. (2011). Invariable biomass-specific primary production of taxonomically discrete picoeukaryote groups across the Atlantic Ocean. *Environ Microbiol* **12**: 3266–3274.
- Hartmann M, Grob C, Tarran GA, Martina AP, Burkill PH, Scanlan DJ *et al.* (2012). Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proc Natl Acad Sci USA* **109**: 5756–5760.
- Hill TCJ, Walsh KA, Harris JA, Mofett BF. (2003). Using ecological diversity measures with bacterial communities. *FEMS Microbiol Ecol* **43**: 1–11.
- Härnström K, Ellegaard M, Andersen TJ, Godhe A. (2011). Hundred years of genetic structure in a sediment revived diatom population. *Proc Natl Acad Sci USA* **108**: 4252–4257.
- Jardillier L, Zubkov MV, Pearman J, Scanlan DJ. (2010). Significant CO<sub>2</sub> fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* **4**: 1180–1192.
- Kirkham AR, Jardillier LE, Holland R, Zubkov MV, Scanlan DJ. (2011b). Analysis of phototrophic picoeukaryote community structure along an extended Ellett Line transect in the northern North Atlantic reveals a dominance of novel prymnesiophyte and prasinophyte phylotypes. *Deep Sea Res I* **58**: 733–744.
- Kirkham AR, Jardillier LE, Tiganescu A, Pearman J, Zubkov MV, Scanlan DJ. (2011a). Basin-scale distribution patterns of photosynthetic picoeukaryotes along an Atlantic Meridional Transect. *Environ Microbiol* **13**: 975–990.
- Le Gall F, Rigaut-Jalabert F, Marie D, Garczareck L, Viprey M, Godet A *et al.* (2008). Picoplankton diversity in the South-East Pacific Ocean from cultures. *Biogeosciences* **5**: 203–214.
- Legendre P, Legendre L. (1998). *Numerical Ecology*, 2nd edn Elsevier Science BV: Amsterdam.
- Lepère C, Vaulot D, Scanlan DJ. (2009). Photosynthetic picoeukaryote community structure in the South East Pacific Ocean encompassing the most oligotrophic waters on Earth. *Environ Microbiol* **11**: 3105–3117.
- Li WKW. (1994). Primary production of prochlorophytes, cyanobacteria, and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39**: 169–175.
- Litchman E, Klausmeier CA. (2008). Trait-based community ecology of phytoplankton. *Annu Rev Ecol Evol Syst* **39**: 615–639.
- Liu H, Probert I, Uits J, Claustre H, Aris-Brosou S, Frada M *et al.* (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proc Natl Acad Sci USA* **106**: 12803–12808.
- Longhurst A. (2007). *Ecological Geography of the Sea*, 2nd edn Elsevier, Academic Press.
- Lozupone C, Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Maguire MJ, Goff LJ, Coleman AW. (1995). In situ plastid and mitochondrial DNA determination: implication of the observed minimal plastid genome number. *Am J Bot* **82**: 1496–1506.
- McDonald SM, Sarno D, Scanlan DJ, Zingone A. (2007). Genetic diversity of eukaryotic ultraphytoplankton in the Gulf of Naples during an annual cycle. *Aquat Microb Ecol* **50**: 75–89.
- Moon van der Staay SY, De Watchter R, Vaulot D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**: 607–610.
- Moore LR, Rocap G, Chisholm SW. (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393**: 464–467.
- Nolte V, Pandey RV, Jost S, Medinger R, Ottenwälder B, Boenigk J *et al.* (2010). Constraining seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol Ecol* **19**: 2908–2915.
- Not F, Latasa M, Scharek R, Viprey M, Karleskind P, Balague V *et al.* (2008). Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. *Deep Sea Res I* **55**: 1456–1473.
- Not F, Massana R, Latasa M, Marie D, Colson C, Eikrem W *et al.* (2005). Late summer community composition and abundance of photosynthetic eukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* **50**: 1677–1686.
- Papke RT, Ward DM. (2004). The importance of physical isolation to microbial diversification. *FEMS Microbiol Ecol* **48**: 293–303.
- Pedros-Alio C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol* **14**: 257–263.
- Pither J. (2007). Comment on ‘Dispersal limitations matter for microbial morphospecies’. *Science* **316**: 1124.
- Richardson TL, Jackson GA. (2007). Small phytoplankton and carbon export from the surface ocean. *Science* **315**: 838–840.
- Rodríguez F, Derelle E, Guillou L, Le Gall F, Vaulot D, Moreau H. (2005). Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ Microbiol* **7**: 853–859.
- Romari K, Vaulot D. (2004). Composition and temporal variability of picoeukaryote community at a coastal site of the English Channel from 18S rDNA sequences. *Limnol Oceanogr* **49**: 784–798.
- Sanders RW, Gast RJ. (2011). Bacteriivory by phototrophic picoplankton and nanoplankton in Arctic waters. *FEMS Microbiol Ecol* **82**: 242–253.
- Schnetzler A, Moorthi SD, Countway PD, Gast RJ, Gilg IC, Caron DA. (2011). Depth matters: Microbial eukaryote diversity and community structure in the eastern North Pacific revealed through environmental gene libraries. *Deep Sea Res I* **58**: 16–26.

- Shi XL, Lepère C, Scanlan DJ, Vault D. (2011). Plastid 16S rRNA gene diversity among eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific Ocean. *PLoS One* **4**: e18979.
- Shi XL, Marie D, Jardillier L, Scanlan DJ, Vault D. (2009). Groups without cultured representatives dominate eukaryotic picophytoplankton in the Oligotrophic South East Pacific. *PLoS One* **4**: e7657.
- Telford RJ, Vandik V, Birks HJB. (2006). Dispersal limitations matter for microbial morphospecies. *Science* **312**: 1015.
- Ter Braak CJF. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* **67**: 1167–1179.
- Tyrell T. (1999). The relative influences of nitrogen and phosphorus on oceanic primary production. *Nature* **400**: 525–531.
- Vault D, Eikrem W, Viprey M, Moreau H. (2008). The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS Microbiol Rev* **32**: 795–820.
- Viprey M, Guillou L, Ferréol M, Vault D. (2008). Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environ Microbiol* **10**: 1804–1822.
- Weber TS, Deutsch C. (2010). Ocean nutrient ratios governed by plankton biogeography. *Nature* **467**: 550–554.
- 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, Schönhuber WA, Fuller NJ, Amann RI, Rippka R, Post AF *et al.* (2001). Closely related *Prochlorococcus* genotypes show remarkably different depth distributions in two oceanic regions as revealed by in situ hybridisation using 16S rRNA-targeted oligonucleotides. *Microbiology* **147**: 1731–1744.
- Worden AZ, Nolan JK, Palenik B. (2004). Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. *Limnol Oceanogr* **49**: 168–179.
- Worden AZ, Not F. (2008). Ecology and diversity of picoeukaryotes. In: Kirchman DL (eds). *Microbial Ecology of the Oceans*, 2nd edn John Wiley & Sons: Hoboken, NJ, USA, pp 159–205.
- Yang J, Smith HG, Sherratt TN, Wilkinson DM. (2010). Is there a size limit for cosmopolitan distribution in free-living microorganisms? A biogeographical analysis of testate amoebae from polar areas. *Microb Ecol* **59**: 635–645.
- Zubkov MV, Fuchs BM, Tarran GA, Burkill PH, Amann R. (2003). High rate of uptake of organic nitrogen compounds by *Prochlorococcus* cyanobacteria as a key to their dominance in oligotrophic oceanic waters. *Appl Environ Microbiol* **69**: 1299–1304.
- Zubkov MV, Tarran GA. (2008). High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature* **455**: 224–226.
- Zwirgmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vault D *et al.* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10**: 147–161.

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