

**Original citation:**

Kirkham, Amy R., Lepère, Cécile, Jardillier, Ludwig E., Not, Fabrice, Bouman, Heather, Mead, A. (Andrew) and Scanlan, David J. . (2013) A global perspective on marine photosynthetic picoeukaryote community structure. The ISME Journal, Vol.7 (No.5). pp. 922-936.

**Permanent WRAP url:**

<http://wrap.warwick.ac.uk/54265>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes the work of researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

© the authors. First published in the ISME Journal published by Nature Publishing Group.

<http://dx.doi.org/10.1038/ismej.2012.166>

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

warwick**publications**wrap  
  
highlight your research

<http://go.warwick.ac.uk/lib-publications>

# A global perspective on marine photosynthetic picoeukaryote community structure

Amy R. Kirkham<sup>1#\*</sup>, Cécile Lepère<sup>1°\*</sup>, Ludwig E. Jardillier<sup>2</sup>, Fabrice Not<sup>3</sup>, Heather  
5 Bouman<sup>4</sup>, Andrew Mead<sup>1</sup> & 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,  
France*

10 <sup>3</sup>*CNRS and Université Pierre et Marie Curie, UMR 7144, Station Biologique de Roscoff,  
Place Georges Teissier, 29680 Roscoff, France.*

<sup>4</sup>*Department of Earth Sciences, University of Oxford, Oxford, UK*

<sup>#</sup>*Present address: School of Environmental Sciences, University of East Anglia, Norwich  
Research Park, Norwich, Norfolk, NR4 7TJ, UK*

15 <sup>°</sup>*Present address: Clermont Université, Université Blaise Pascal, Laboratoire  
"Microorganismes: Génome et Environnement", BP 10448, F-63000 Clermont-Ferrand*

<sup>\*</sup>These authors contributed equally to this work

20 Correspondence: D.J. Scanlan, School of Life Sciences, University of Warwick, Gibbet Hill  
Road, Coventry, CV4 7AL, UK. Email: [d.j.scanlan@warwick.ac.uk](mailto:d.j.scanlan@warwick.ac.uk)

## Summary

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) 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 (inter-basins) 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 to assess their evolution and contribution to CO<sub>2</sub> fixation especially in the context of global climate change.

## Introduction

Understanding the mechanisms that generate and maintain biodiversity, and more particularly  
45 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 macro-organism  
biogeography, with only relatively few recent studies for micro-organisms (e.g. see Foissner  
*et al.*, 2006; Fuhrman *et al.*, 2008; Caron, 2009; Nolte *et al.*, 2010). Opposing views have  
50 claimed that free-living microbial taxa present a cosmopolitan distribution on the one hand,  
(Finlay & Fenchel, 2004, Finlay *et al.*, 2006; Pither, 2007) with particular implications for  
organisms in the smallest size classes (e.g.  $<0.5 \mu\text{m}$ ) (Yang *et al.*, 2010), or on the other are  
composed primarily of species that have limited geographical distributions (Papke & Ward,  
2004; Telford *et al.*, 2006; Hähnström *et al.* 2011).

55 During the last decade, molecular techniques have begun to provide insights into the  
structure of small ( $<3 \mu\text{m}$  diameter), unicellular photosynthetic picoeukaryote (PPE)  
communities in various marine environments (reviewed by Vaultot *et al.*, 2008). In particular,  
cloning and sequencing of both nuclear (e.g. see Moon-van der Staay *et al.*, 2001; Not *et al.*,  
2008) and plastid (e.g. Fuller *et al.*, 2006b; MacDonald *et al.*, 2007; Lepère *et al.*, 2009)  
60 small subunit rRNA genes has revealed a vast, previously unsuspected diversity within this  
assemblage. This focus is largely due to 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; Jardillier *et al.*, 2010;  
Cuvelier *et al.*, 2010; Grob *et al.*, 2011). Moreover, PPEs have recently been shown to be key  
65 bacterivores controlling bacterioplankton abundance in both temperate (Zubkov and Tarran,  
2008) and oligotrophic gyre (Hartmann *et al.*, 2012) waters. Furthermore, oceanic N:P ratios

are likely to be governed by plankton biogeography, due to 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  
70 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, four major ocean biomes) by different molecular methods, targeting the gene encoding the plastid 16S rRNA.

## 75 **Experimental Procedures**

### *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) summarized in Table 1. The cruise transects encompass i) all four major ocean biomes  
80 (Longhurst 2007): polar, westerly winds, trade winds, and coastal boundary domains ii) various temperature zones: arctic, temperate, subtropical and tropical and iii) 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  
85 cruises (Figure 1, Table 1) can be found in Not *et al.* (2005), Not *et al.* (2008) and Bouman *et al.* (2006). In brief, samples analyzed from the Arctic Ocean cruise were collected at six depths from the surface of the water column to approximately 60 m deep in August 2002. Indian Ocean samples were collected from the upper 800m on the VANC10MV cruise during May-June 2003 between Cape Town (South Africa) and Port Hedland (Australia), passing

90 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. Due to 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  
95 bottles. For DNA extraction, 10 L of seawater was filtered first through a 47-mm diameter, 3  
µm pore size polycarbonate filter (Millipore MCE MF) and then onto a 47-mm diameter, 0.45  
µm pore size polysulfone filter (Supor450, Gelman Sciences, AnnArbor, Mich) 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, 50 mM Tris pH 9.0), flash-  
100 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),  
Kirkham *et al.* (2011a and b). We also include sequencing data from a time-series taken from  
the Gulf of Naples, Mediterranean Sea (see McDonald *et al.*, 2007). Environmental  
105 parameters 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,  
Beckton Dickinson, Oxford, UK) using their characteristic pigment autofluorescence and  
size. The flow rate was calculated by adding a known concentration of 0.5 µm multi-  
110 fluorescent latex beads (Polysciences, Eppelheim, Germany) as an internal standard. Flow  
cytometry data were processed using CellQuest software (Beckton Dickinson, Oxford, UK).

#### *Polymerase Chain Reaction (PCR) conditions*

115 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 approximately 830 bp PCR product.

#### *Dot blot hybridisation analysis*

120 16S rDNA amplicons from environmental DNAs and control strains were purified, blotted onto nylon membranes and hybridized 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, 125 Pinguiphyceae, Prasinophyceae clade VI (Prasinococcales), Prymnesiophyceae and Trebouxiophyceae, respectively (Fuller *et al.*, 2006). 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, 130 <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). Hybridization was quantified by using a Fujifilm FLA-5000 phosphorimager and Total Laboratory software (Phoretix). The relative hybridisation of a given specific probe compared with that of the eubacterial probe to the control DNAs was 135 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) and screened by restriction fragment length polymorphism (RFLP) after digestion with *HaeIII* and *EcoRI* as described previously (Kirkham *et al.*, 2011a).  
140

### *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.*, (2006), Lepère *et al.*, (2009); Shi *et al.*, (2011), Kirkham *et al.*, (2011a), McDonald *et al.*, (2007) and Kirkham *et al.*, (2011b), respectively.  
145

The Margalef index (Hill *et al.*, 2003) was used for quantifying the richness (alpha-diversity) of PPEs since, 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.  
150  
155

To explore relationships between environmental parameters and the distribution of PPEs measured by hybridization 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  
160



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 amongst the samples, constrained to be maximally related to weighted linear combinations of the environmental variables (Legendre & 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 & Legendre, 1998). Pairwise 2-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, 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  $\text{ml}^{-1}$  encountered in the Rockall trough of the extended Ellett Line transect (Figure 2, Supplementary Table 1). The lowest abundances were found in the Arabian Sea transect (mean over the transect  $8.2 \times 10^2$  cells  $\text{ml}^{-1}$ ), being below detection limits for much of the

185 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  $\text{ml}^{-1}$  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  
190 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 (e.g. Zubkov *et al.*, 2003). Although there is a general trend in the PPE/total  
195 picophytoplankton (i.e. (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 due to a multifactorial effect of greater biovolume,  
200 higher growth rates, and high grazing mortality rates (Li, 1994; Worden *et al.* 2004; Jardillier  
*et al.*, 2010; Grob *et al.*, 2011).

### ***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  
205 species richness values (alpha diversity) using the Margalef index ( $D_{mg}$ ) (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  
( $D_{mg} = 7.3$ ). This high richness was positively correlated with temperature ( $r = 0.84$ ,  $p < 0.05$ )

and low nutrient concentrations ( $r = 0.71$ ,  $p < 0.05$ ) i.e. 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 analyzed 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 analyzed and on average comprise 78% of the total relative hybridisation values obtained with the ten PPE class-specific probes used over all seven 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 (FISH) studies confirm the high abundance of these pico-prymnesiophytes in the Atlantic (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 (Jardillier *et al.*, 2010; Cuvelier *et al.*, 2010) even though they mostly lack cultured representatives. The success of these organisms could be due to mixotrophic behaviour (i.e. 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 which 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).  
260 Their low abundance in open ocean waters has also been observed using FISH analysis (e.g. 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 (e.g. Romari and Vaultot 2004). Moreover, cultured  
265 cryptophytes are generally larger than 5µm though 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% or more 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  
270 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  
275 detected in over 80% of the samples analysed, including the entire Arctic Ocean and BEAGLE cruises. This may be due to 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, whilst

280 Jardillier *et al.* (2010) reported that PPEs of a size  $1.8\pm 0.1$   $\mu\text{m}$  comprised 14-57% Pelagophyceae via FISH analysis across a range of samples in the subtropical North Atlantic.

*PPE class distributions: relationships to environmental variables*

Canonical correspondence analyses (CCA) was performed on each cruise separately, revealing associations between PPE class distributions and physico-chemical 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 % change in composition explained by the measured environmental variables was on average 50%. CCAs were also performed for the global dataset (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 eleven 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  
305 coordinates on the first two canonical axes from the overall CCA (Fig. 5), higher abundance  
of Chrysophyceae is associated with samples 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 PO<sub>4</sub> concentrations. Conversely, higher abundance  
of Prymnesiophyceae is associated with samples taken later in the year, at higher latitudes,  
310 higher NO<sub>3</sub>+NO<sub>2</sub> 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 PO<sub>4</sub> and  
higher N:P ratios. In contrast, Prasinophyceae, Trebouxiophyceae, Eustigmatophyceae and  
Chrysophyceae are associated with samples collected in the early spring, with higher  
315 concentrations of PO<sub>4</sub> 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 XXXX ( $P < 0.01$ ). Prymnesiophyceae had a  
correlation coefficient of -0.3 with spring ( $P < 0.01$ ), 0.21 with latitude ( $P < 0.01$ ), 0.2 with  
320 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.

Considering the variables of phosphate and NO<sub>3</sub>+NO<sub>2</sub> concentrations in more detail, peaks in  
relative hybridisation values for the Prymnesiophyceae tended to be linked with lower PO<sub>4</sub>  
325 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 PO<sub>4</sub> 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 were calculated to be 14:1 by linear regression, similar to the study of Tyrell (1999) that reported a 15:1 N:P ratio. Considering only samples for which Prymnesiophyceae relative hybridisations were greater than 30%, the average N:P ratio was 25:1, whereas considering only samples for which Chrysophyceae relative hybridisation values were greater than 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 due to 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 (e.g. see Shi *et al.*, 2011; Moon van der Staay *et al.*, 2001) 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; Supplementary Table 2). *P*-value matrices using the total PPE dataset 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 to 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 since libraries of similar

status are well spread on the tree, e.g. 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  
380 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 0.28-1.37 mg m<sup>-3</sup>) 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  
385 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,  
390 (EEG3, EIB4), which are geographically adjacent stations of comparable trophic and temperature status present a very similar overall composition (Figure 6a; Supplementary Table 2). However, their Prymnesiophyceae sequences are genetically distinct (Supplementary Table 2b). Specific lineages observed within the Prymnesiophyceae (e.g. Lepère *et al.*, 2009) likely include various ecotypes subject to different long- and short-term  
395 ecological constraints (Worden and Not, 2008). In this respect, metagenomic analysis of a natural pico-prymnesiophyte population revealed a mosaic gene repertoire including specific adaptations for growth in oligotrophic environments (Cuvelier *et al.*, 2010), conditions which 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;

400 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 Table 2c,e). This may suggest that these classes contain much less variation  
405 than the Prymnesiophyceae and Prasinophyceae (Supplementary Table 2b,d) and may instead  
contain widespread lineages (Figure 6b,c) with generalist growth strategies rather than  
differently niche-adapted lineages.

### *Conclusions*

410 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 even though the  
alpha diversity of PPEs decrease with depth their community structure (beta diversity) did not  
415 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 whilst  
differentiated microbial communities exist along an environmental gradient (Green and  
Bohannan, 2006). Although the sampling was not exhaustive, and many more taxa (especially  
420 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 (inter-basins) 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  
425 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, e.g. those encompassing high and low light-  
adapted ecotypes (Rodriguez *et al.*, 2005), likely occur (e.g. the prasinophyte *Ostreococcus*  
(Demir- Hilton *et al.*, 2011). Further study of PPE distribution patterns at a higher taxonomic  
430 resolution is thus required and renewed efforts are needed to obtain into culture many PPE  
representatives only known from environmental gene sequencing.

## Acknowledgements

The authors would like to 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  
435 award to ARK and NERC grant NE/C003160/1 to D.J.S.

## References

- Andersen RA. (2004). Biology and systematics of heterokont and haptophyte algae. *Am J Botany* **10**: 1508-22.
- 440 Arrigo KR, Robinson DH, Worthen DL, Dunbar RB, DiTullio GR, VanWoert M and Lizotte, MP. (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, Ulloa O, Scanlan DJ, Zwirgmaier K, Li WK, Platt T, Stuart V, Barlow R, Leth  
445 O, Clementson L, Lutz V, Fukasawa M, Watanabe S, Sathyendranath S. (2006). Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312**: 918-21.
- Bouman HA, Lepère C, Scanlan DJ, Ulloa O. (2012). Phytoplankton community structure in a high-nutrient, low-chlorophyll region of the Southern Ocean during winter-mixed  
450 and summer-stratified conditions. *Deep Sea Res I* (*in press*)
- Caron DA. (2009). Past President's Address: Protistan Biogeography: Why All The Fuss? *J Euk Microbiol* **56**: 105-112.
- Cuvelier ML, Allen A, Monier A, McCrow JP, Messié M, Tringe SG, Woyke T, Welsh RM, Ishoey T, Lee J-H, Binder BJ, DuPont CL, Latasa M, Guigand C, Buck KR, Hilton J,  
455 Thiagarajan M, Caler E, Read B, Lasken RS, Chavez FP, Worden A. (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 and Worden AZ. (2011). Global distribution patterns of distinct clades of the photosynthetic picoeukaryote

- 460 *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 *Coccolithophores – from molecular processes to global impact*, ed. H. Thierstein, JR Young, pp. 429-53. Amsterdam: Elsevier.
- Finlay BJ, Fenchel T. (2004). Cosmopolitan metapopulations of free-living microbial  
465 eukaryotes. *Protist* **155**: 237–244.
- Finlay BJ, Esteban GF, Brown S, Fenchel T, Hoef-Emden K. (2006). Multiple cosmopolitan ecotypes within a microbial eukaryote morphospecies. *Protist* **157**: 377-390.
- Foissner, W. (2006). Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozool* **45**: 111-136.
- 470 Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL, Brown JH. (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* **105**: 7774-7778.
- Fuller NJ, Marie D, Partensky F, Vaultot D, Post AF, Scanlan DJ. (2003). Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine  
475 *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, Le Gall F, Scanlan DJ. (2006a). Molecular analysis of photosynthetic picoeukaryote community structure along an Arabian Sea transect. *Limnol Oceanogr* **51**: 2502-  
480 2514.
- Fuller NJ, Campbell C, Allen DJ, Pitt FD, Zwirgmaier K, Le Gall F, Vaultot D, Scanlan DJ. (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.
- 485 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**: 3266-74.
- 490 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.
- Hartmann M, Grob C, Tarran GA, Martina AP, Burkill PH , Scanlan DJ, and Zubkov MV. (2012). Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proc Natl Acad Sci USA* **109**: 5756-5760.
- 495 Hill TCJ, Walsh KA, Harris JA, Mofett BF. (2003). Using ecological diversity measures with bacterial communities. *FEMS Microbiol Ecol* **43**: 1-11.
- 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.
- 500 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.
- Kirkham AR, Jardillier LE, Holland R, Zubkov MV, Scanlan DJ. (2011b). Analysis of phototsynthetic picoeukaryote community structure along an extended Ellett Line
- 505



- transect in the northern North Atlantic reveals a dominance of novel prymnesiophyte and prasinophyte phylotypes. *Deep Sea Res I* **58**: 733-744.
- Le Gall F, Rigaut-Jalabert F, Marie D, Garczareck L, Viprey M, Godet A, Vaulot D. (2008). Picoplankton diversity in the South-East Pacific Ocean from cultured. *Biogeosciences* **5**: 203-214.
- 510
- Legendre P, Legendre L. (1998). Numerical Ecology. (2<sup>nd</sup> edition). 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.
- 515
- 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.
- 520
- Liu H, Probert I, Uits J, Claustre H, Aris-Brosou S, Frada M, Not F, de Vargas C. (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. 2<sup>nd</sup> edn. Elsevier, Academic Press.
- 525
- Lozupone C, Knight R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228-35.
- 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  
530 determination: Implication of the observed minimal plastid genome number. *Am J Botany* **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.
- 535 Moon van der Staay SY, De Watchter R, Vaultot 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, Ottenwalder B, Boenigk J, Schlotterer C. (2010).  
540 Constrasting 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, Ontoriaoviedo I, Cumino A,  
Goetze E, Vaultot D. (2008). Protistan assemblages across the Indian Ocean, with a  
specific emphasis on the picoeukaryotes. *Deep Sea Res I* **55**: 1456-1473.
- 545 Not F, Massana R, Latasa M, Marie D, Colson C, Eikrem W, Pedros-Alio C, Vaultot D, Simon  
N. (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.
- 550 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 and Jackson GA. (2007). Small phytoplankton and carbon export from the  
555 surface ocean. *Science* **315**: 838-840.
- Rodríguez F, Derelle E, Guillou L, Le Gall F, Vaultot D, Moreau H. (2005). Ecotype diversity  
in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ  
Microbiol* **7**: 853-859.
- Romari K, Vaultot D. (2004). Composition and temporal variability of picoeukaryote  
560 community at a coastal site of the English Channel from 18S rDNA sequences.  
*Limnol Oceanogr* **49**: 784-798.
- Sanders RW, Gast RJ. (2011). Bacterivory by phototrophic picoplankton and nanoplankton in  
Arctic waters. *FEMS Microbiol Ecol* **82**: 242-253.
- Schnetzer A, Moorthi SD, Countway PD, Gast RJ, Gilg IC, Caron DA. (2011). Depth  
565 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, Marie D, Jardillier L, Scanlan DJ, Vaultot D. (2009). Groups without cultured  
representatives dominate eukaryotic picophytoplankton in the Oligotrophic South  
East Pacific. *PLoS One* **4**: e7657.
- 570 Shi XL, Lepère C, Scanlan DJ, Vaultot D. (2011). Plastid 16S rRNA gene diversity among  
eukaryotic picophytoplankton sorted by flow cytometry from the South Pacific  
Ocean. *PLoS One* **4**: e18979.
- Telford RJ, Vandik V, Birks HJB. (2006). Dispersal limitations matter for microbial  
morphospecies. *Science* **312**: 1015.

- 575 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.
- Vaulot D, Eikrem W, Viprey M, Moreau H. (2008). The diversity of small eukaryotic  
580 phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS Microbiol Rev* **32**: 795-820.
- 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*  
585 **65**: 2585-2591.
- West NJ, Schönhuber 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 hybridisation using 16S  
rRNA-targeted oligonucleotides. *Microbiology* **147**: 1731–1744.
- 590 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 (ed.).  
Microbial Ecology of the Oceans. John Wiley & Sons, Hoboken, NJ, USA, Second  
595 Edition, 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  
600 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, Vaultot D, Not F, Massana  
605 R, Ulloa O, Scanlan DJ. (2008). Global phylogeography of marine *Synechococcus*  
and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic  
biomes. *Environ Microbiol* **10**: 147-161.

## Figure and Table legends

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

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

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

615 **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 3 CTD's from where DNA was pooled. Contour plots (a, b) indicate the  
620 % 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.

**Figure 5:** Canonical correspondence analysis plot using relative hybridisation values (%)  
625 detected for all cruises except the Indian Ocean transect. PPE classes are as follows: Prymnesiophyceae (Prym), Chrysophyceae (Chry), Cryptophyceae (Cryp), Pinguiphyceae (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 (PO<sub>4</sub>), nitrate + nitrite  
630 concentration (NO<sub>2</sub>NO<sub>3</sub>), 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, (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 ( $\geq 25^\circ\text{C}$ ).

Table 1: Details of cruises analysed in this study

Table 2: PPE alpha-diversity calculated using the Margalef index (DMg). 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 ( $\geq 25^\circ\text{C}$ ).

## Supplementary Information

**Supplementary Table 1:** Environmental parameters measured on the seven cruises analysed: depth, salinity, temperature, chlorophyll *a* (Chl *a*), dissolved oxygen (Diss Ox),  
655 concentrations of NO<sub>2</sub>, NO<sub>3</sub>, NH<sub>4</sub>, PO<sub>4</sub>, SiO<sub>4</sub>, and abundance of *Prochlorococcus*,  
*Synechococcus* and PPEs. (x 10<sup>3</sup> cells ml<sup>-1</sup>). <sup>CL</sup> Samples where clone libraries have been  
constructed; \* not measured.

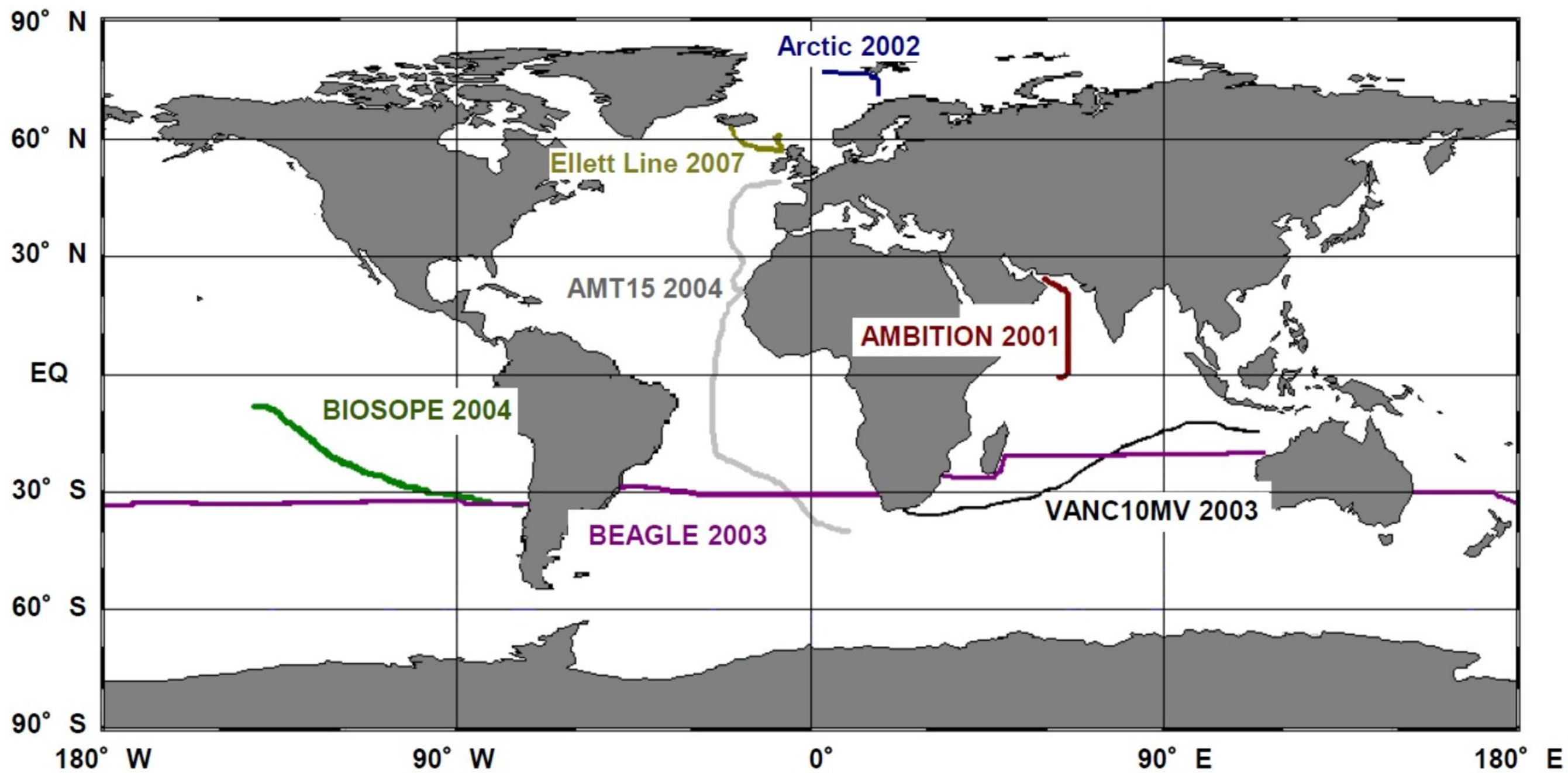
**Supplementary Table 2:** Unifrac analysis matrix showing the *P*-values when comparing  
660 each clone library (16S rRNA gene sequences) to each other, for (a) the total PPE community  
(b), Prymnesiophyceae (c) Chrysophyceae (d) Prasinophyceae and (e) Cryptophyceae. Values  
in bold are significant (p<0.05).

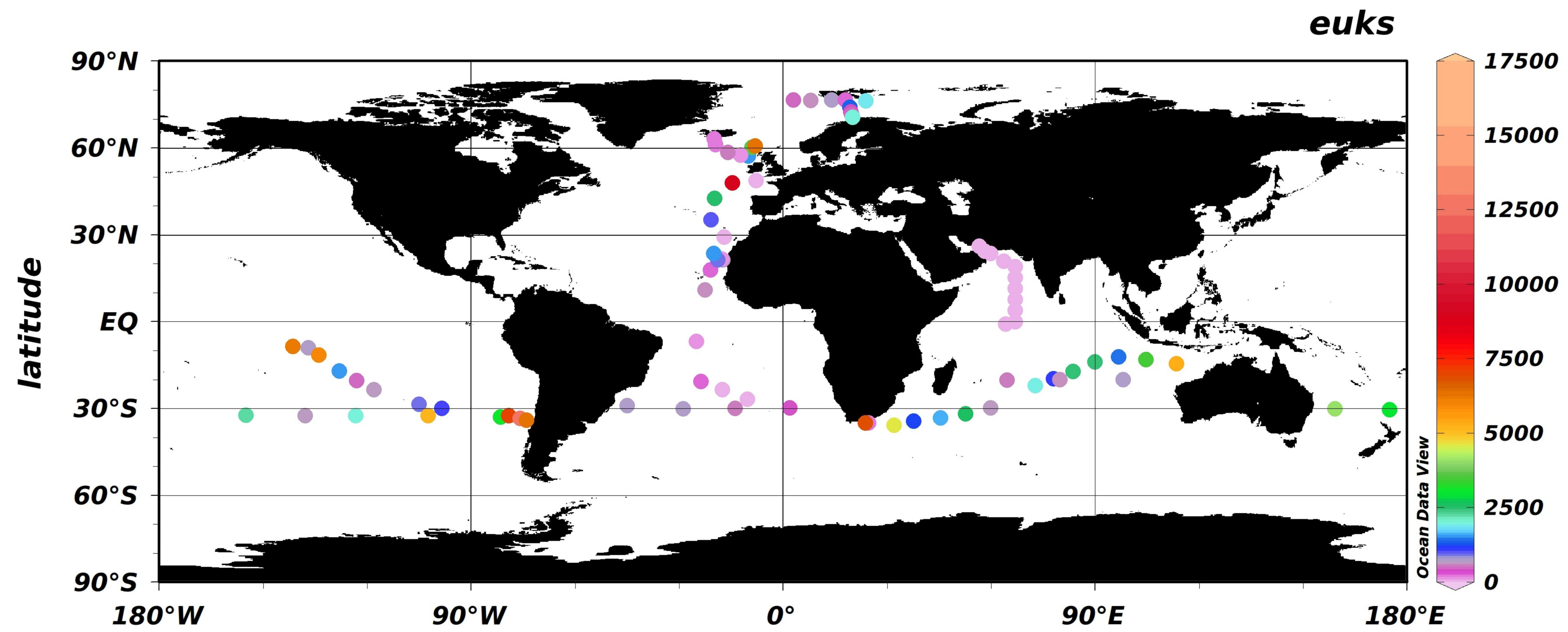
**Supplementary Table 3:** Percent relative hybridisation values obtained by class-specific dot-  
665 blot hybridisations for each of the seven cruises analysed. PPE classes are abbreviated as  
follows: Prymnesiophyceae (Prym), Chrysophyceae (Chry), Cryptophyceae (Cryp),  
Pinguiphyceae (Ping), Pelagophyceae (Pela), Eustigmatophyceae (Eust) and  
Trebouxiophyceae (Treb). (\*= not determined).

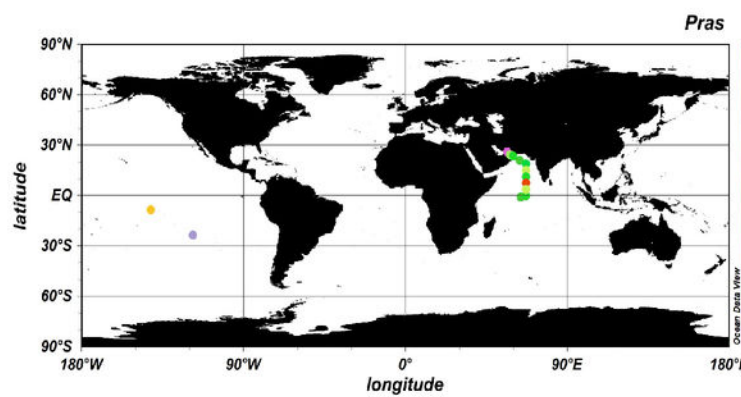
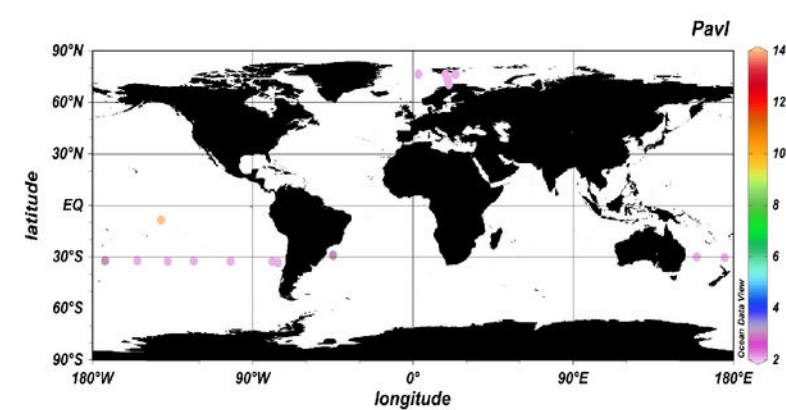
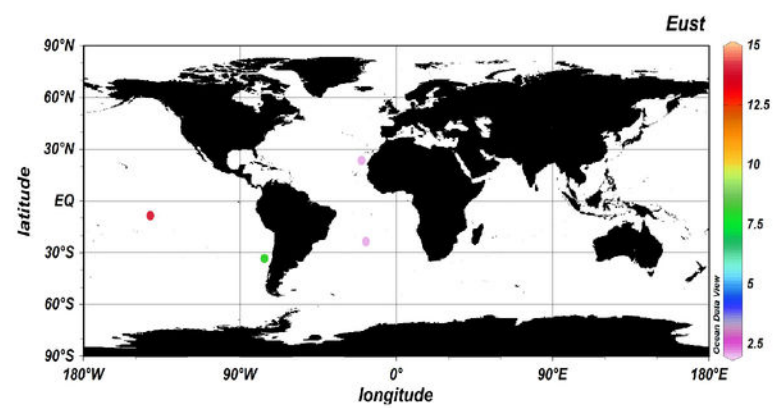
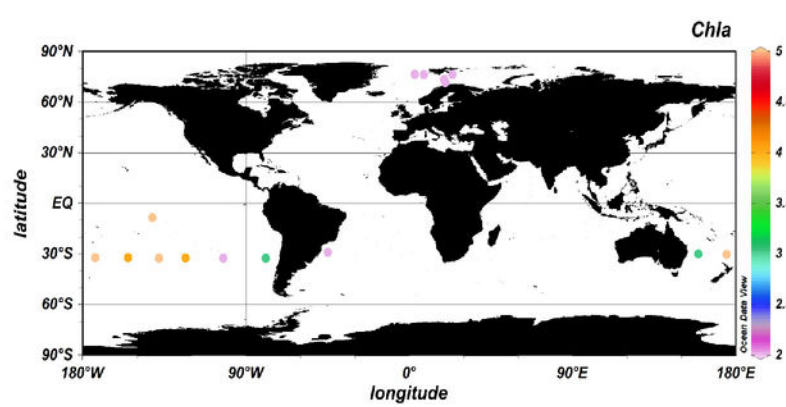
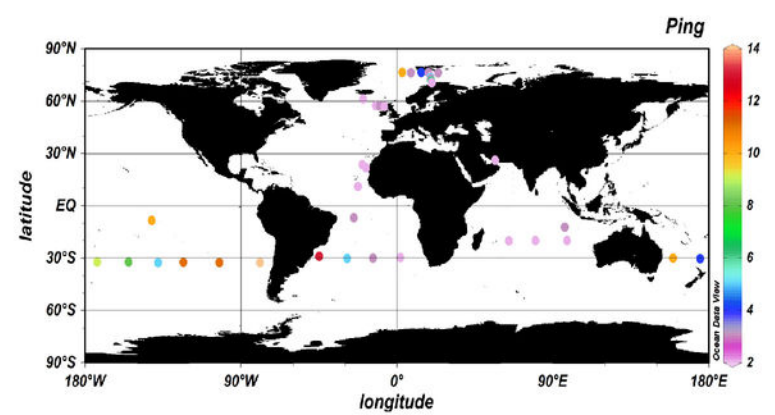
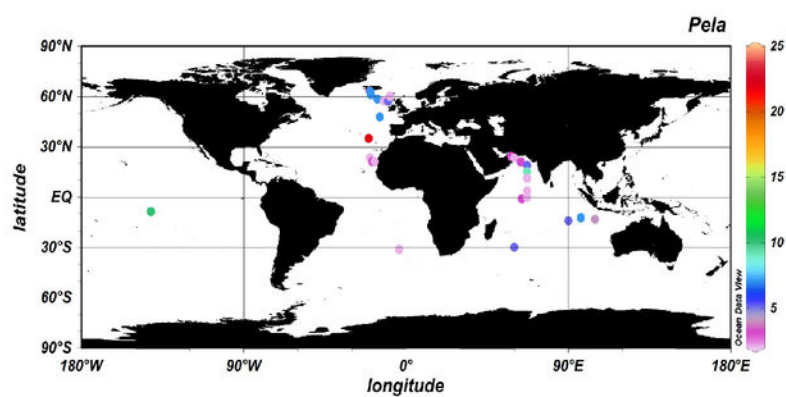
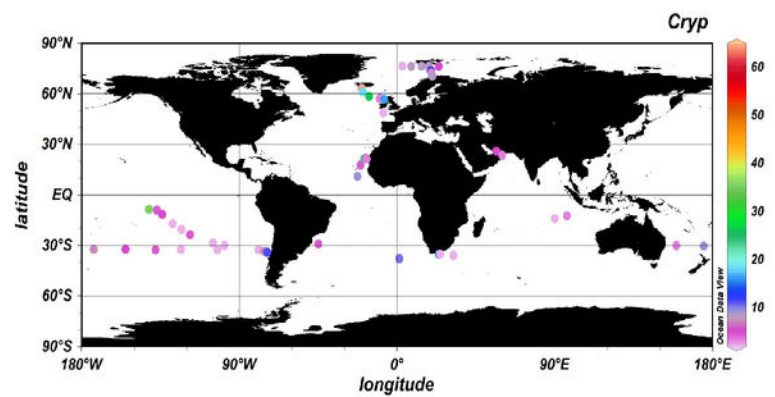
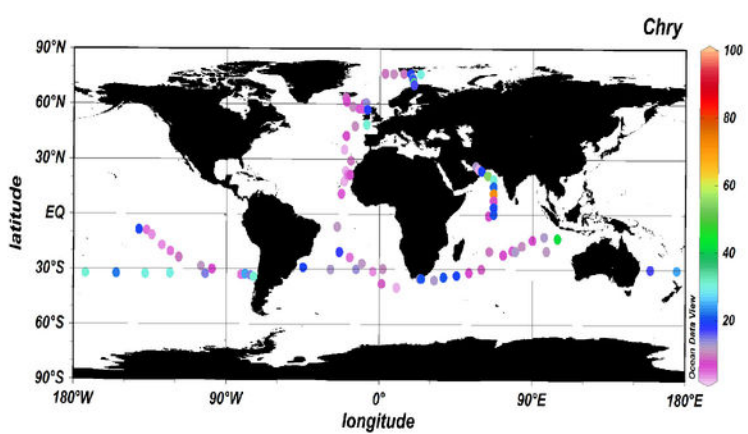
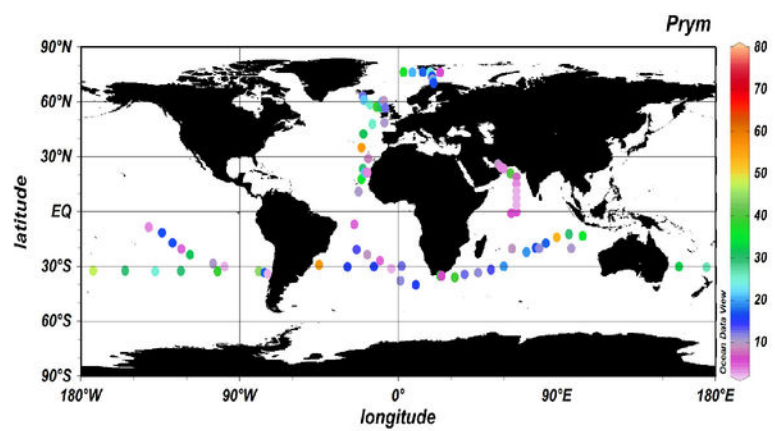
670 **Supplementary Figure 1:** Canonical correspondence analysis plot using relative  
hybridisation values (%) hybridisation detected for (a) the Arctic transect and (b) the  
BEAGLE transect. PPE classes are as follows: Prymnesiophyceae (Prym), Chrysophyceae  
(Chry), Cryptophyceae (Cryp), Pinguiphyceae (Ping), Pelagophyceae (Pela),  
Eustigmatophyceae (Eust), Chlorarachniophyceae (Chlo) and Trebouxiophyceae (Treb).

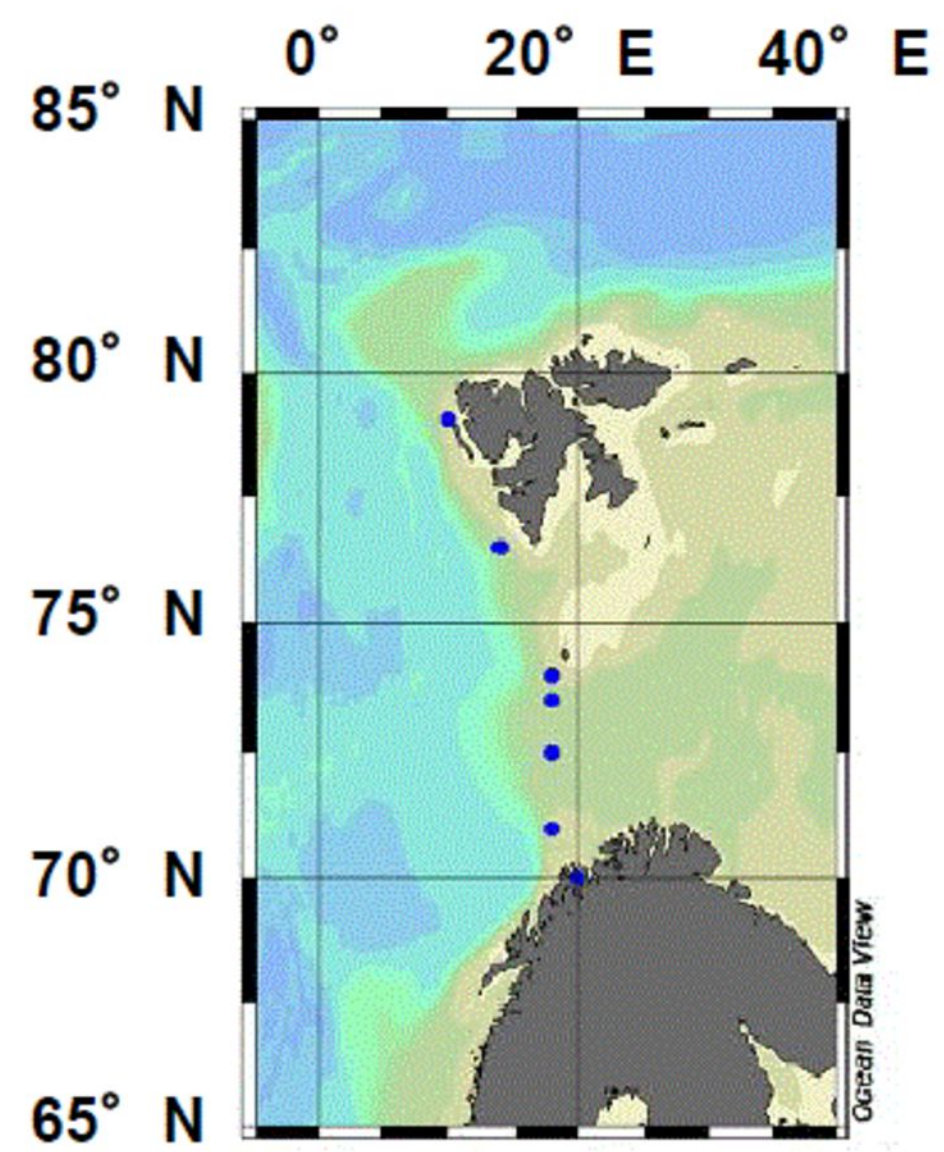
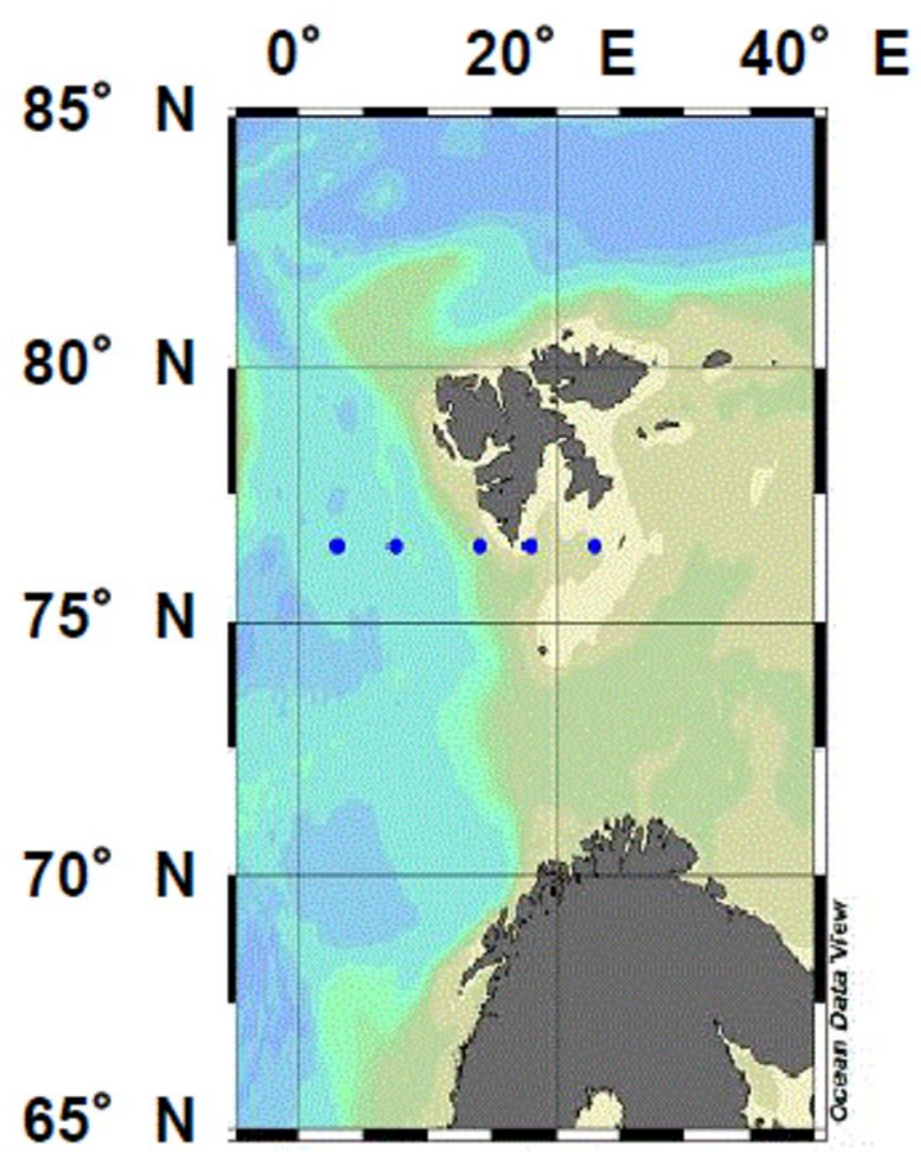


675 Variables are spring (Spring), chlorophyll *a* (Chl), phosphate (PO<sub>4</sub>), nitrate + nitrite  
concentration (NO<sub>3</sub>+NO<sub>2</sub>), nitrate + nitrite:phosphate ratio (N:P), salinity (Sal), depth  
(Depth), mixed layer depth (MLD), photosynthetically active radiation level averaged over  
the mixed layer (PARML), latitude (Lat) and temperature (Temp). The x axis for the Arctic  
680 transect explains 62.3% and the y axis explains 12.1% of the variation in dot blot  
hybridisation data. The x axis for the BEAGLE transect explains 50.4% and the y axis  
explains 38.7%.

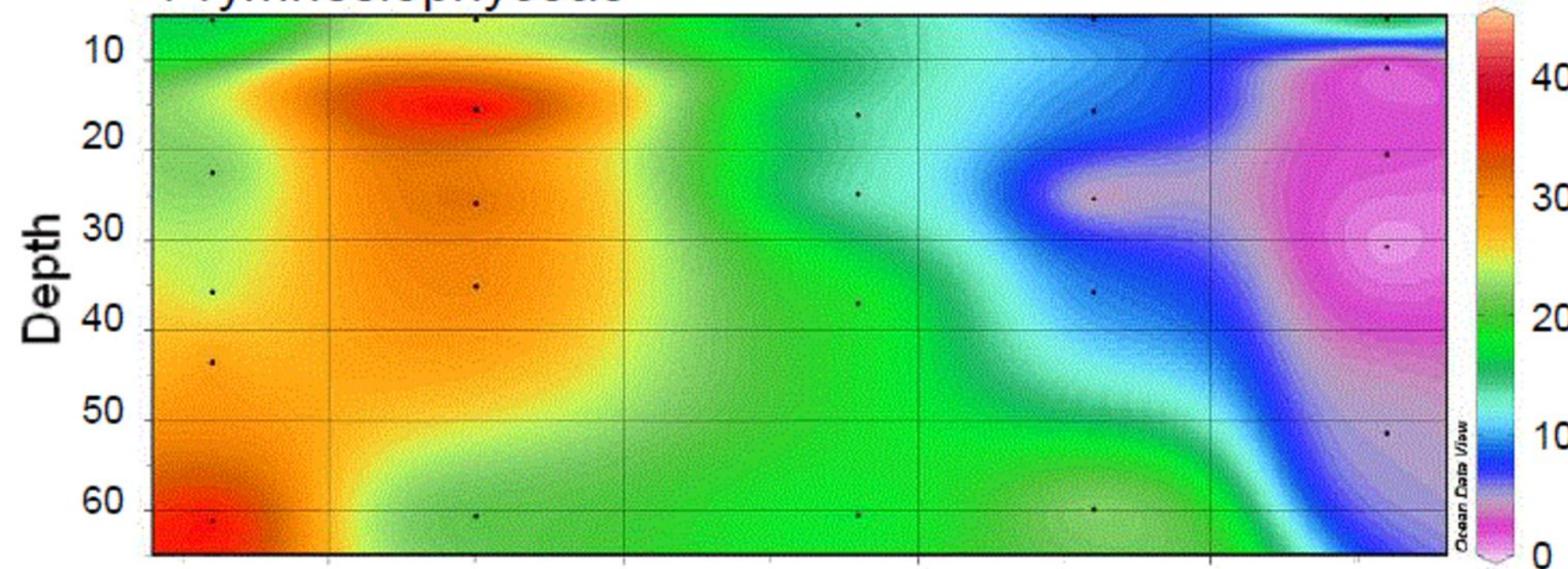




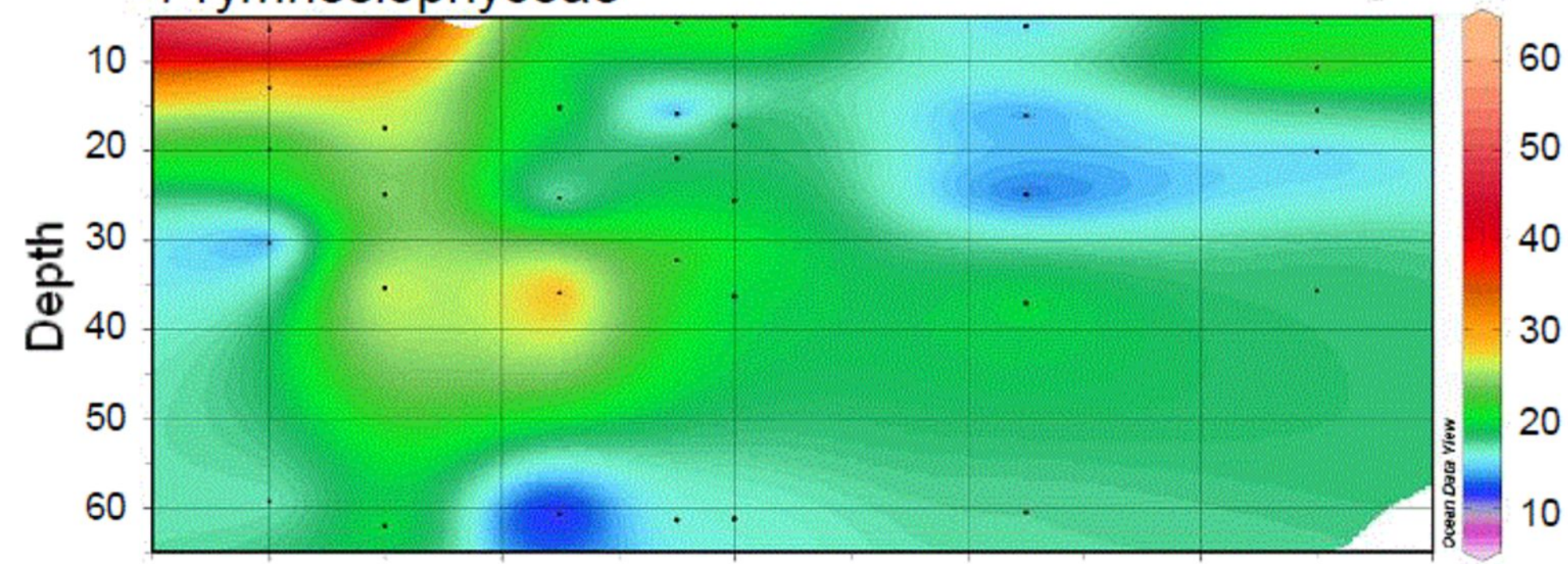




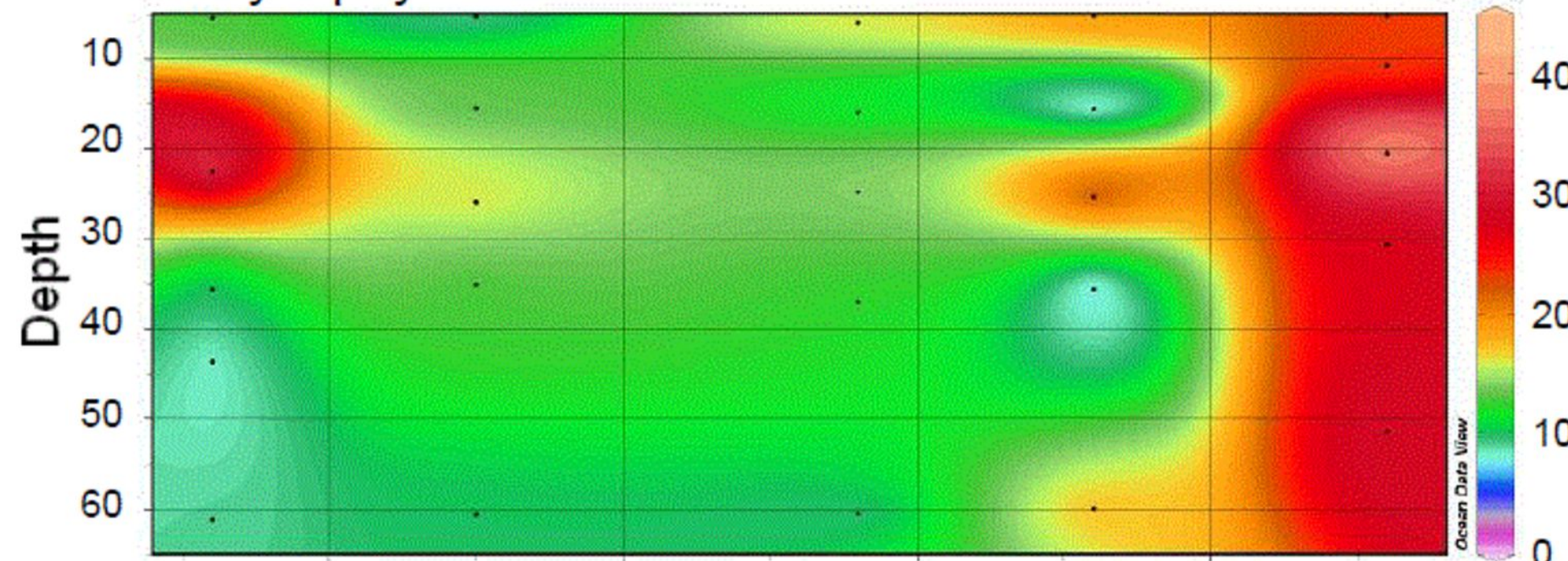
Prymnesiophyceae



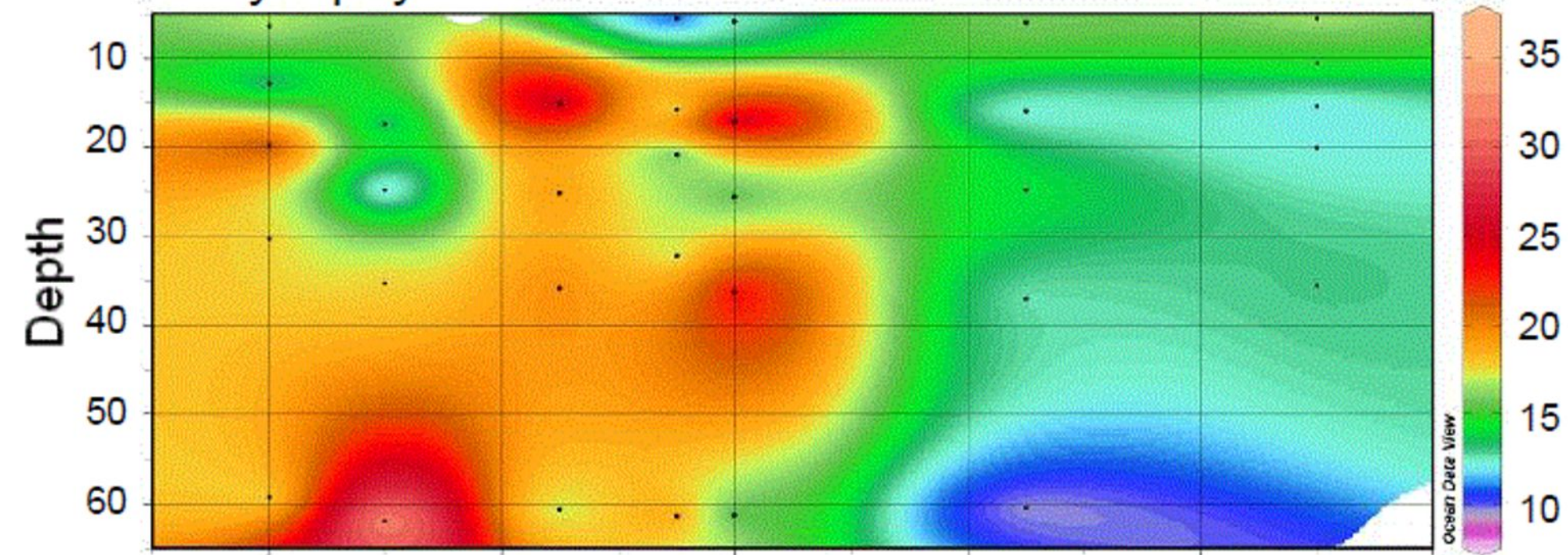
Prymnesiophyceae



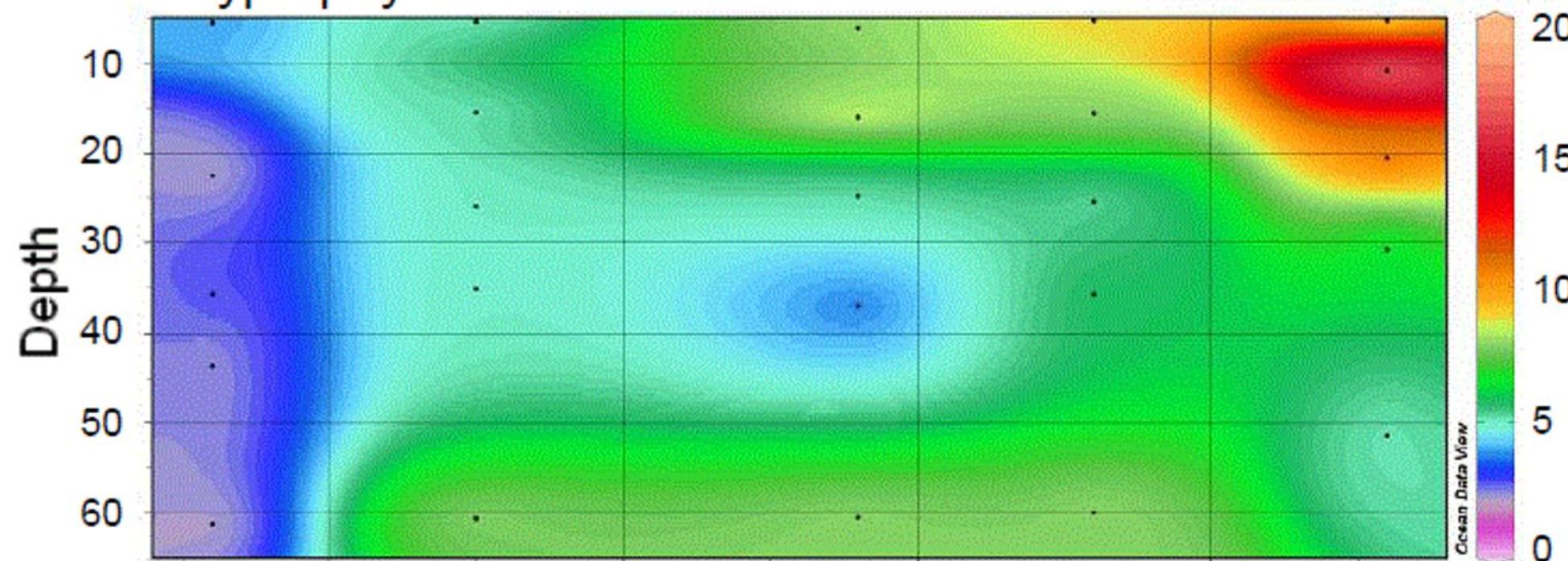
Chrysophyceae



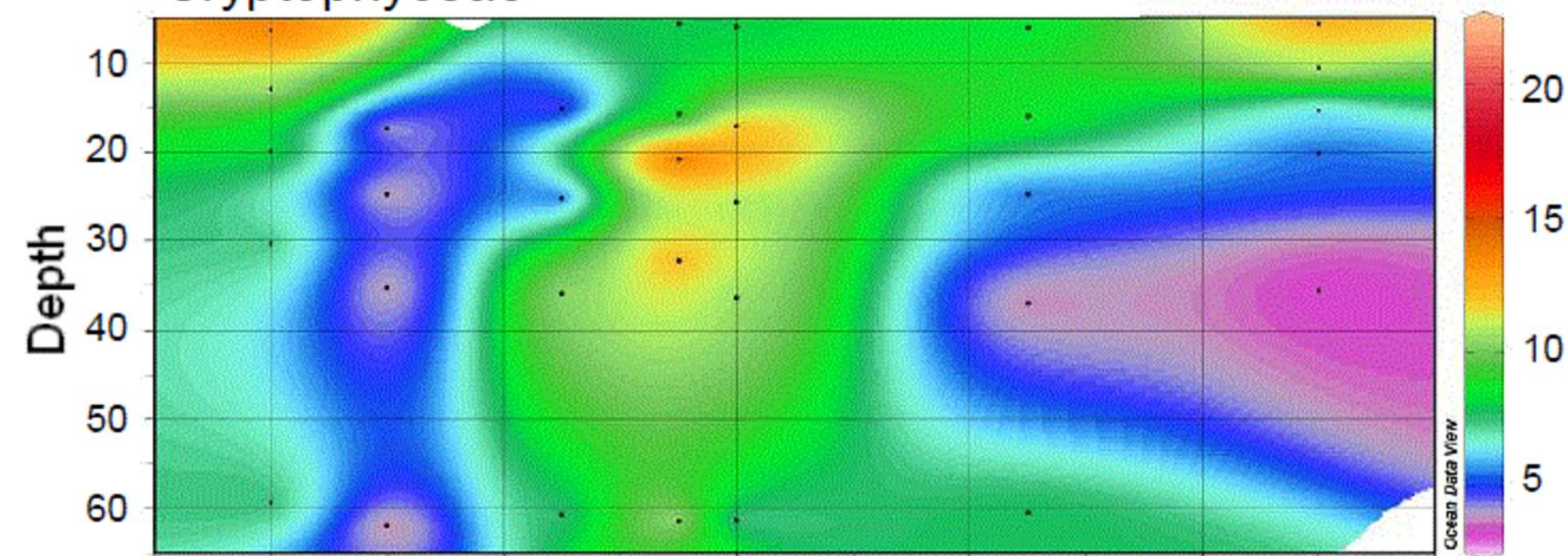
Chrysophyceae



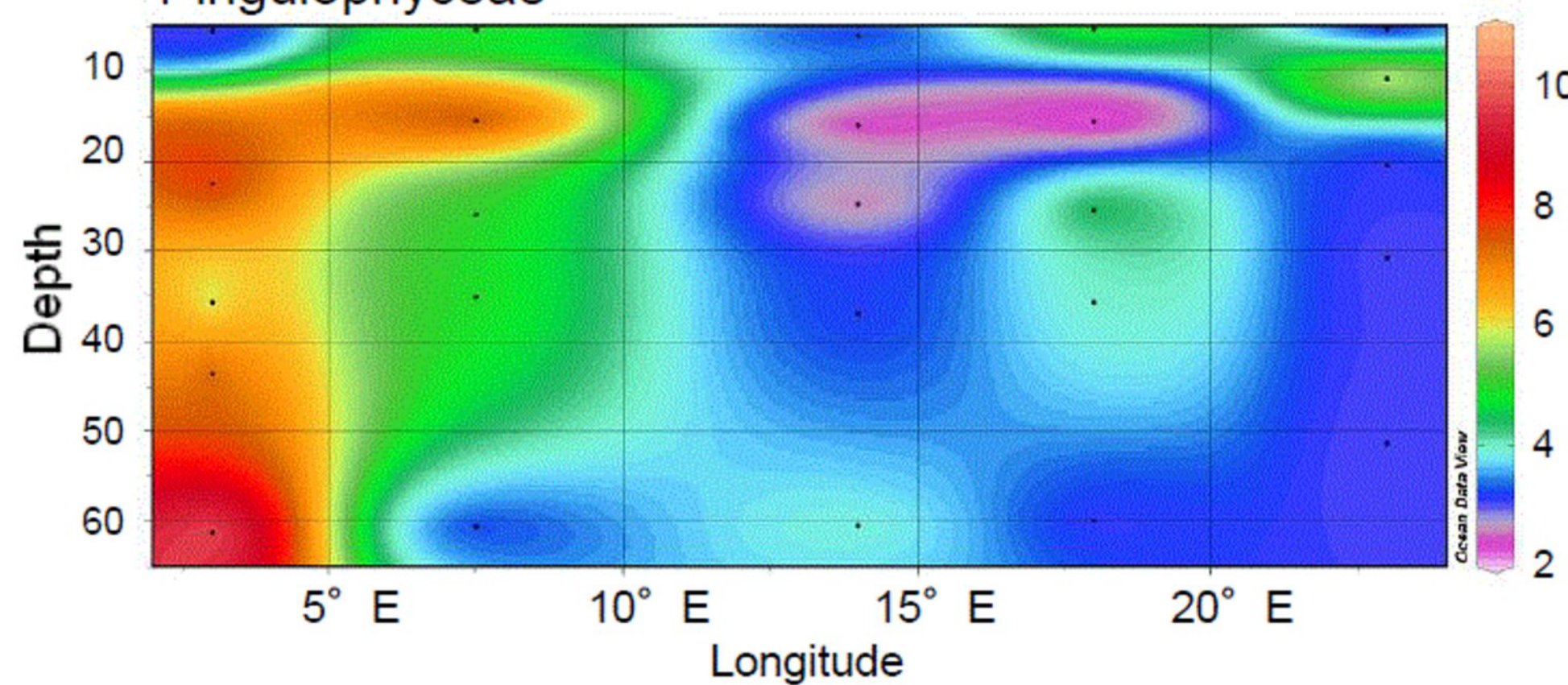
Cryptophyceae



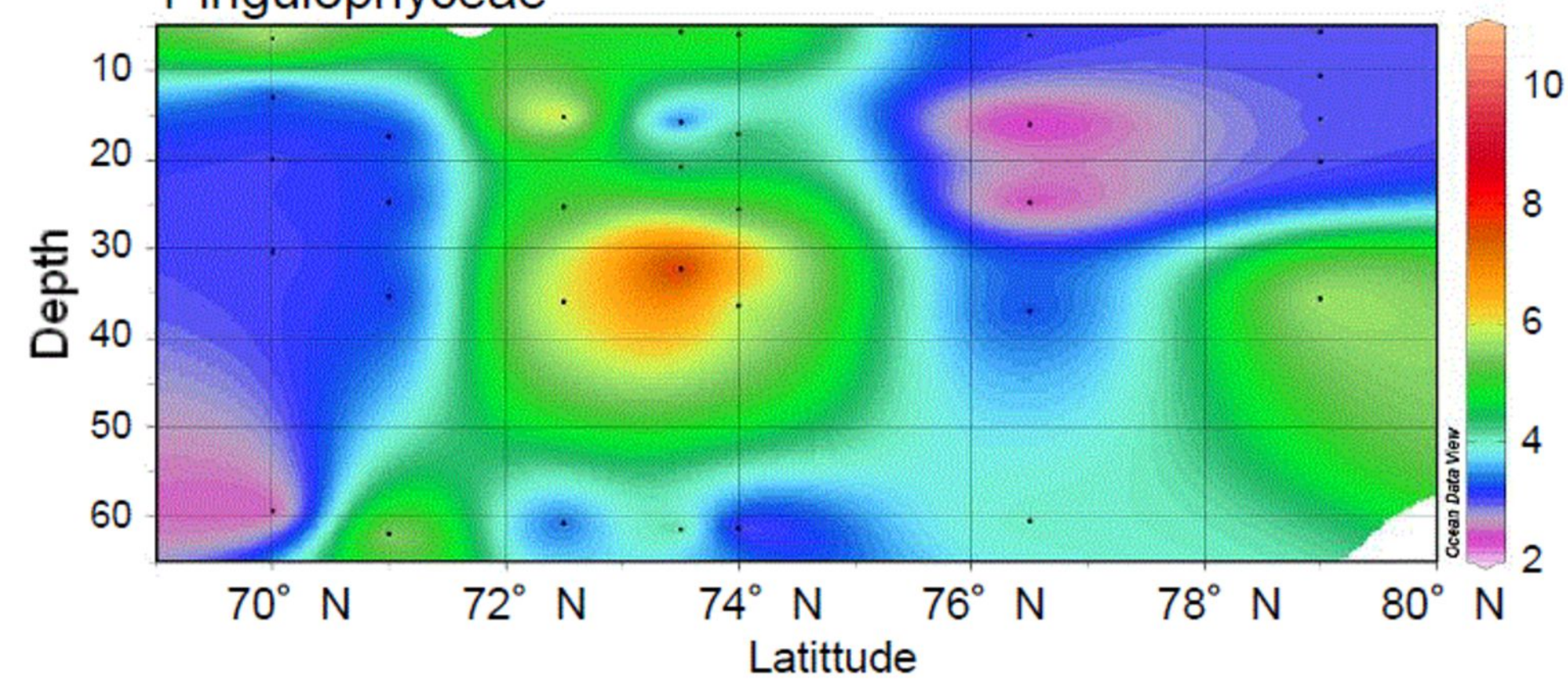
Cryptophyceae

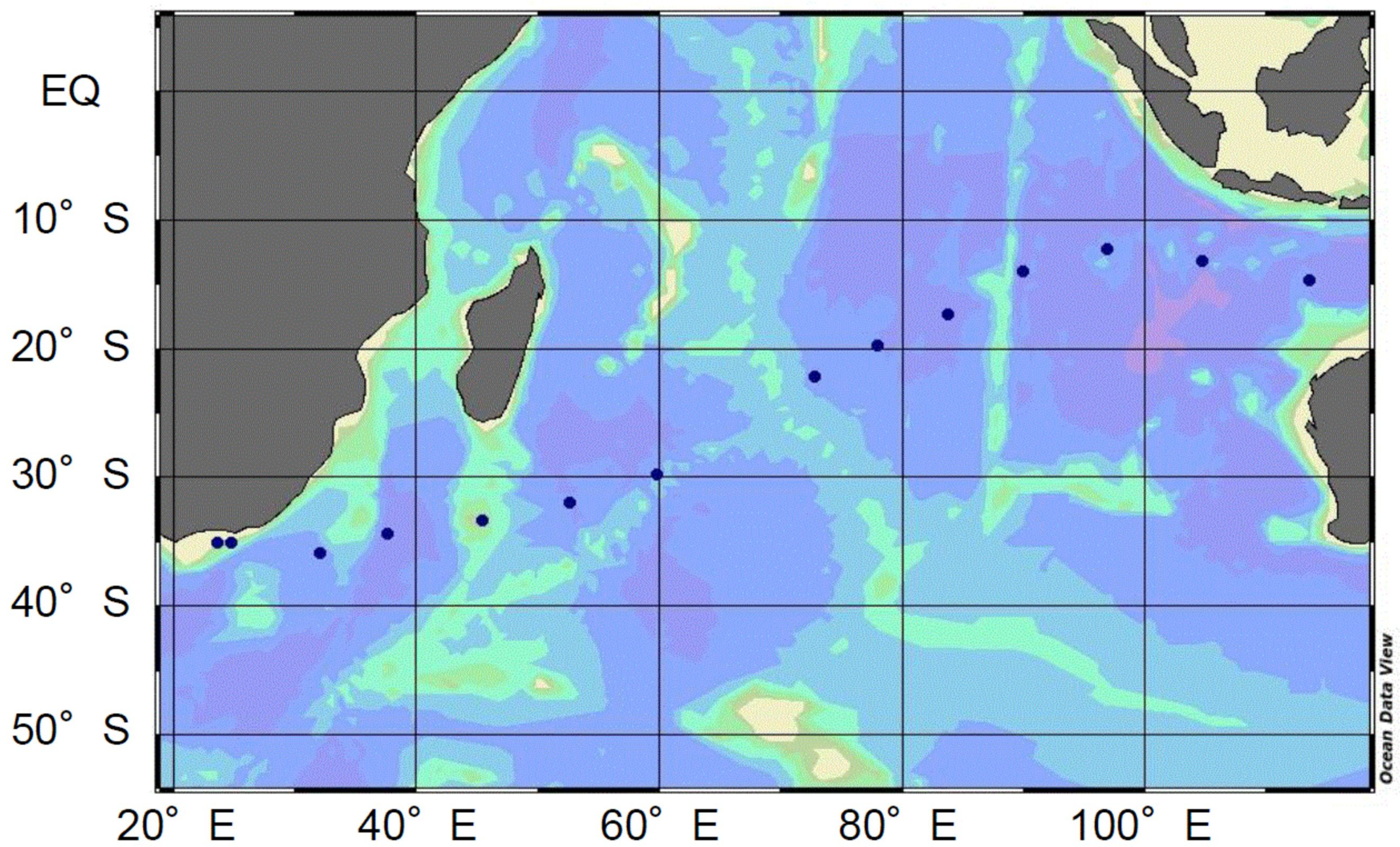


Pinguiphyceae

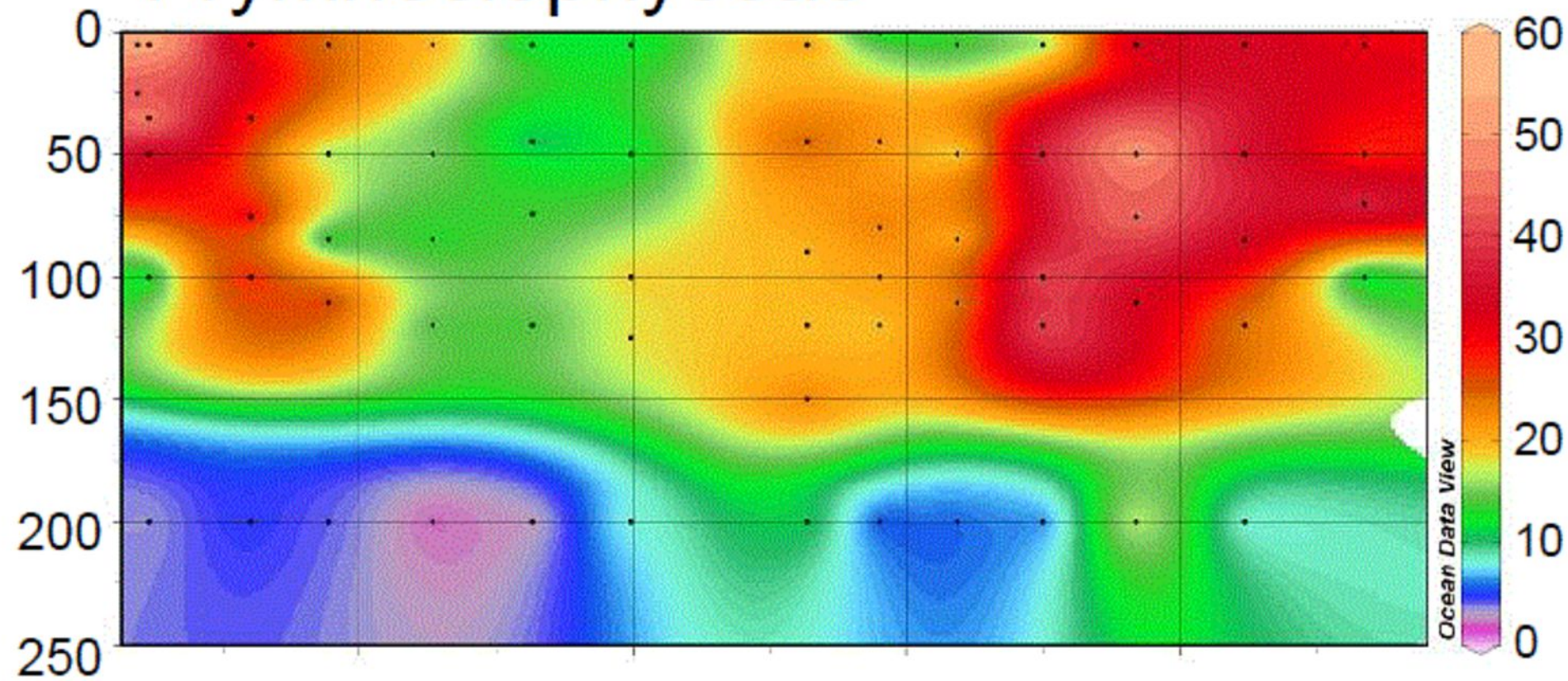


Pinguiphyceae

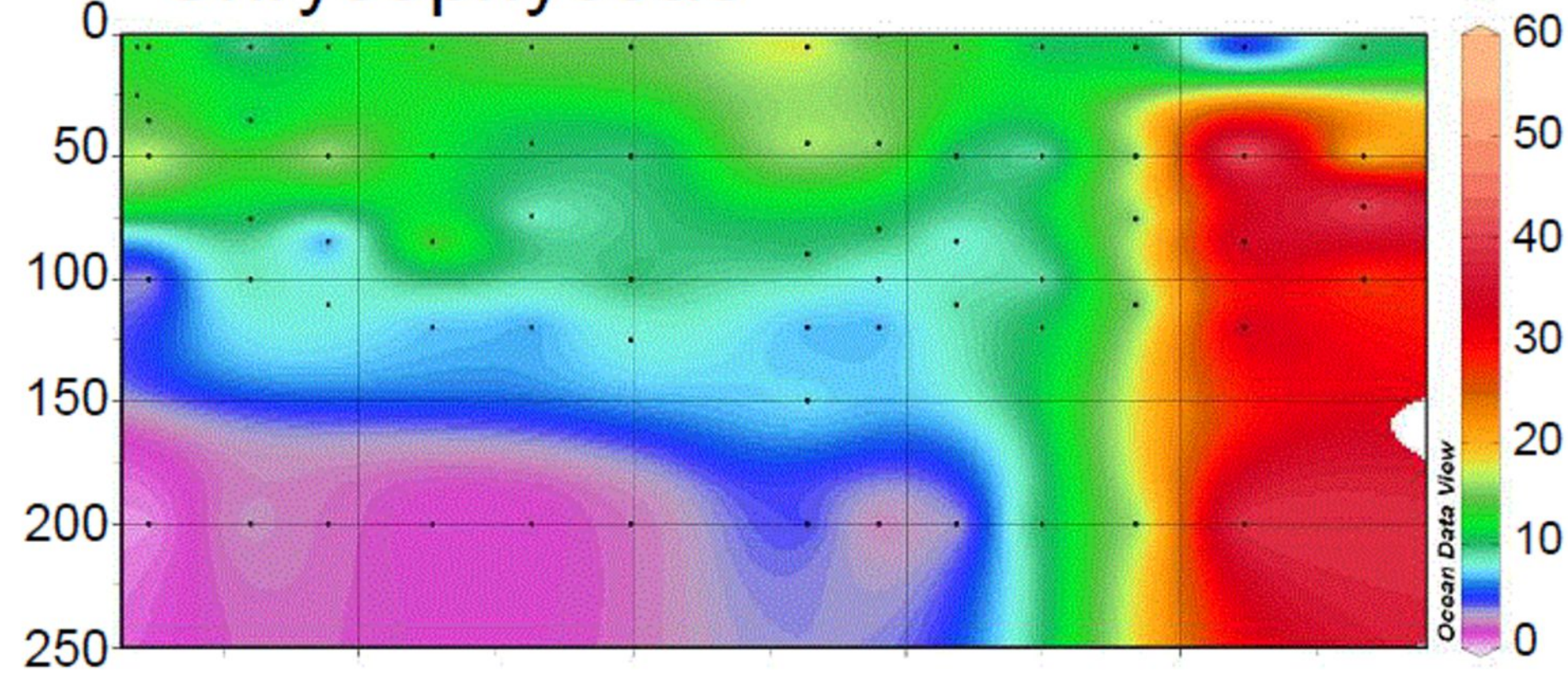




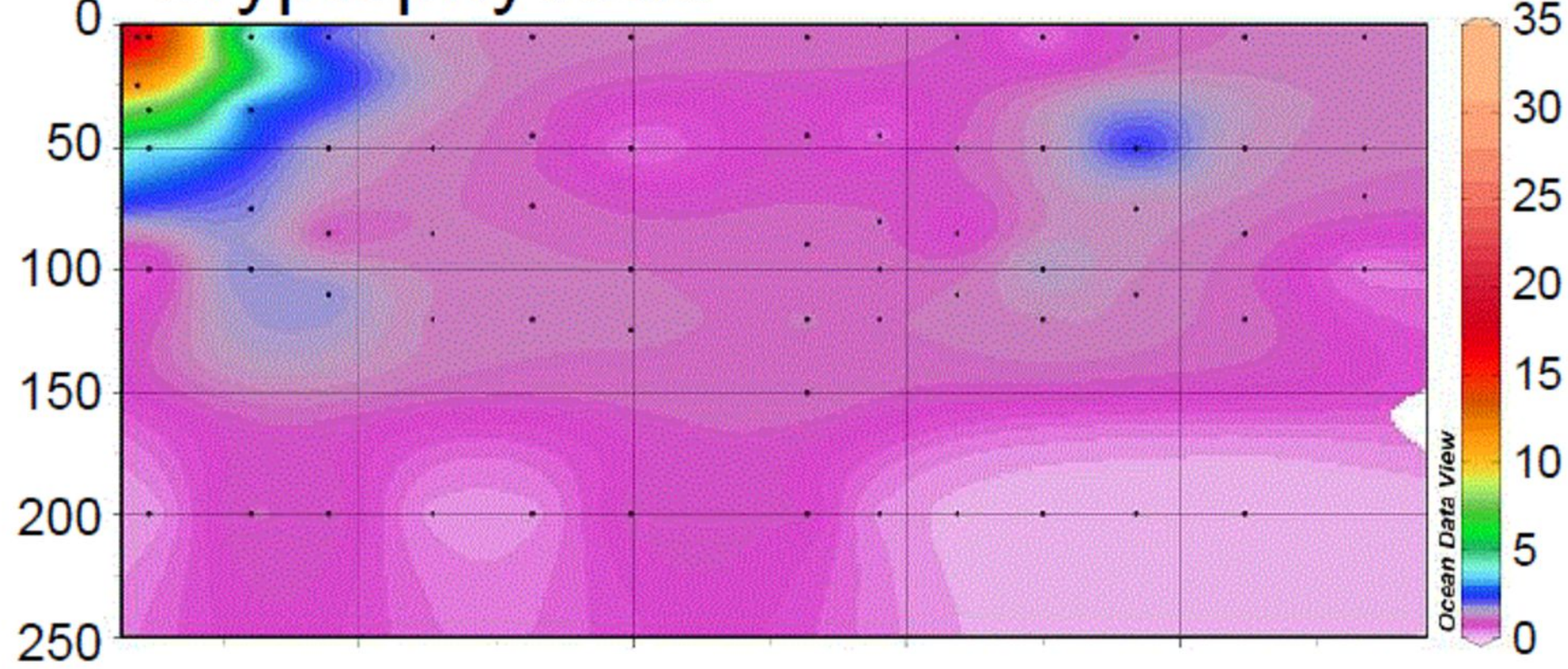
Prymnesiophyceae



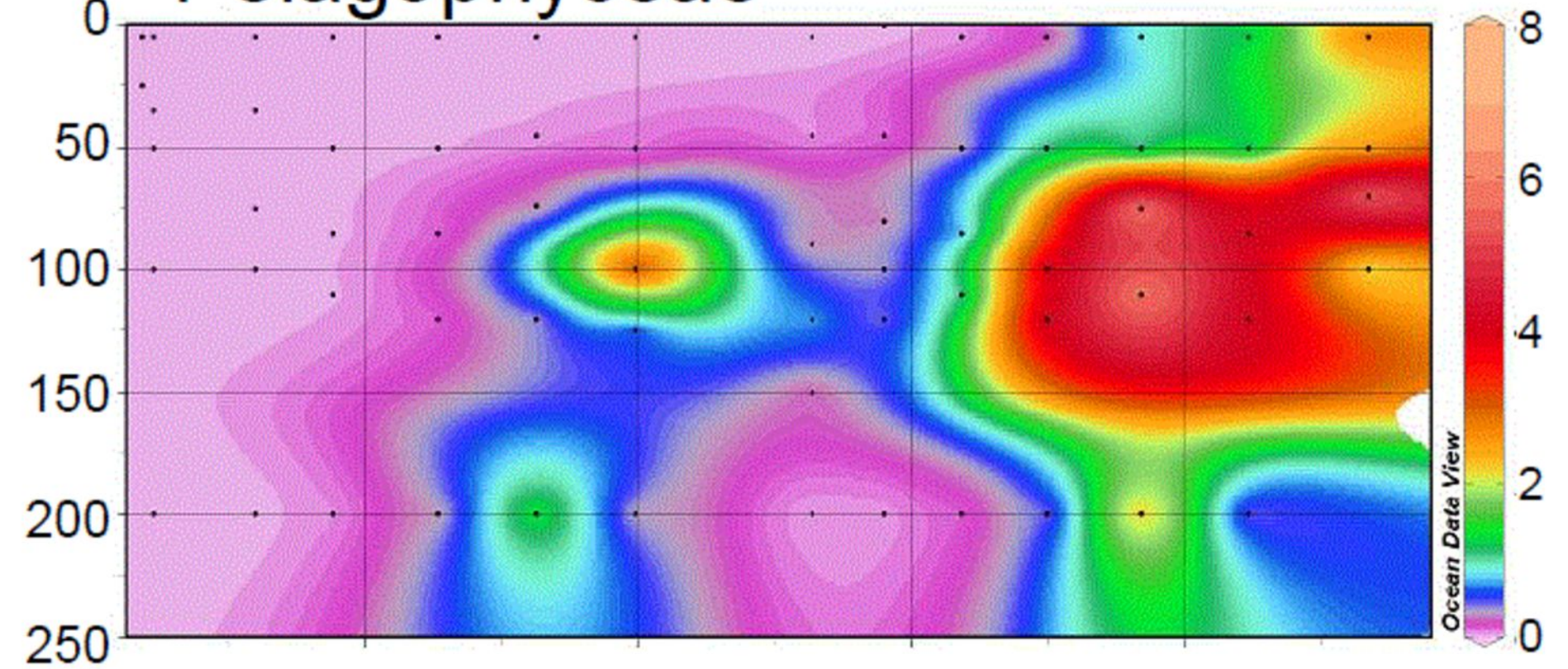
Chrysophyceae



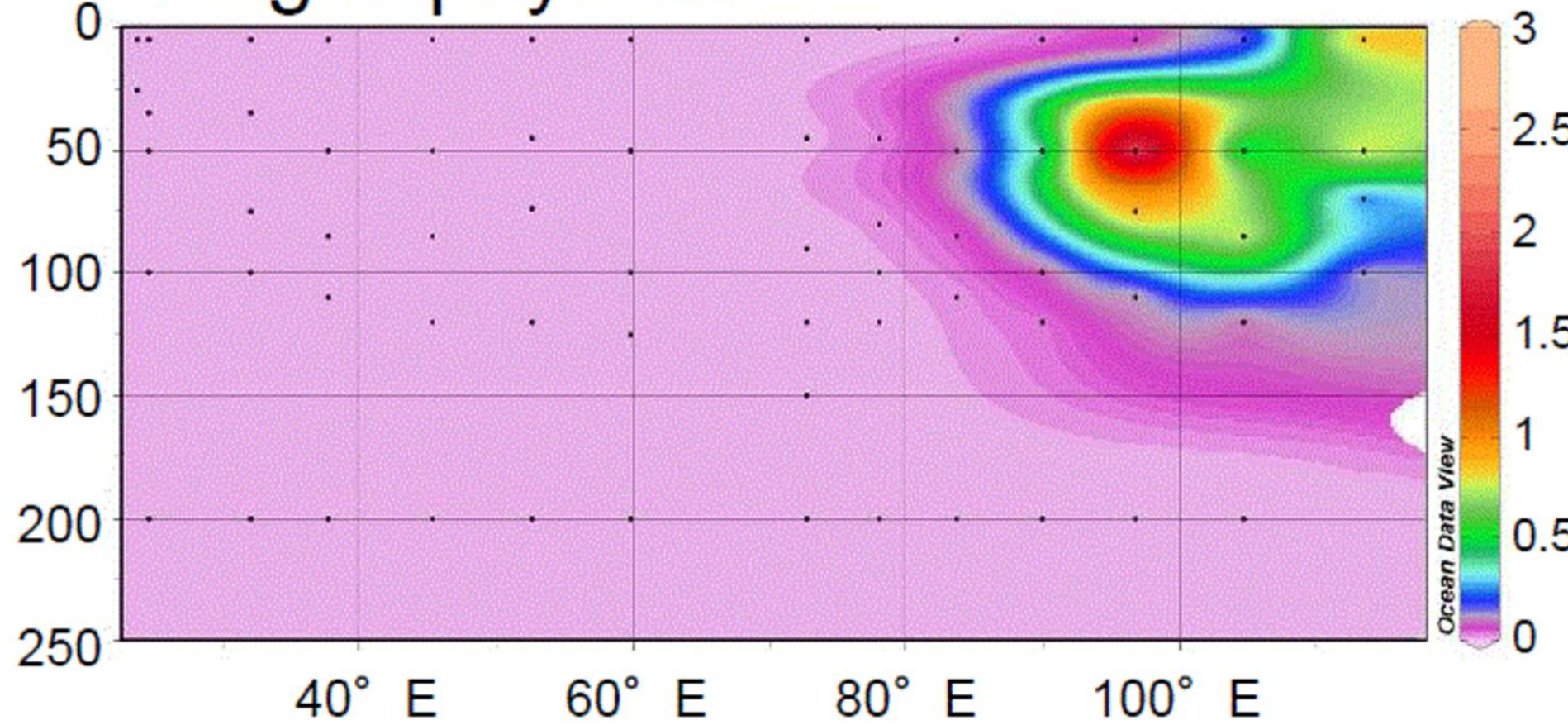
Cryptophyceae



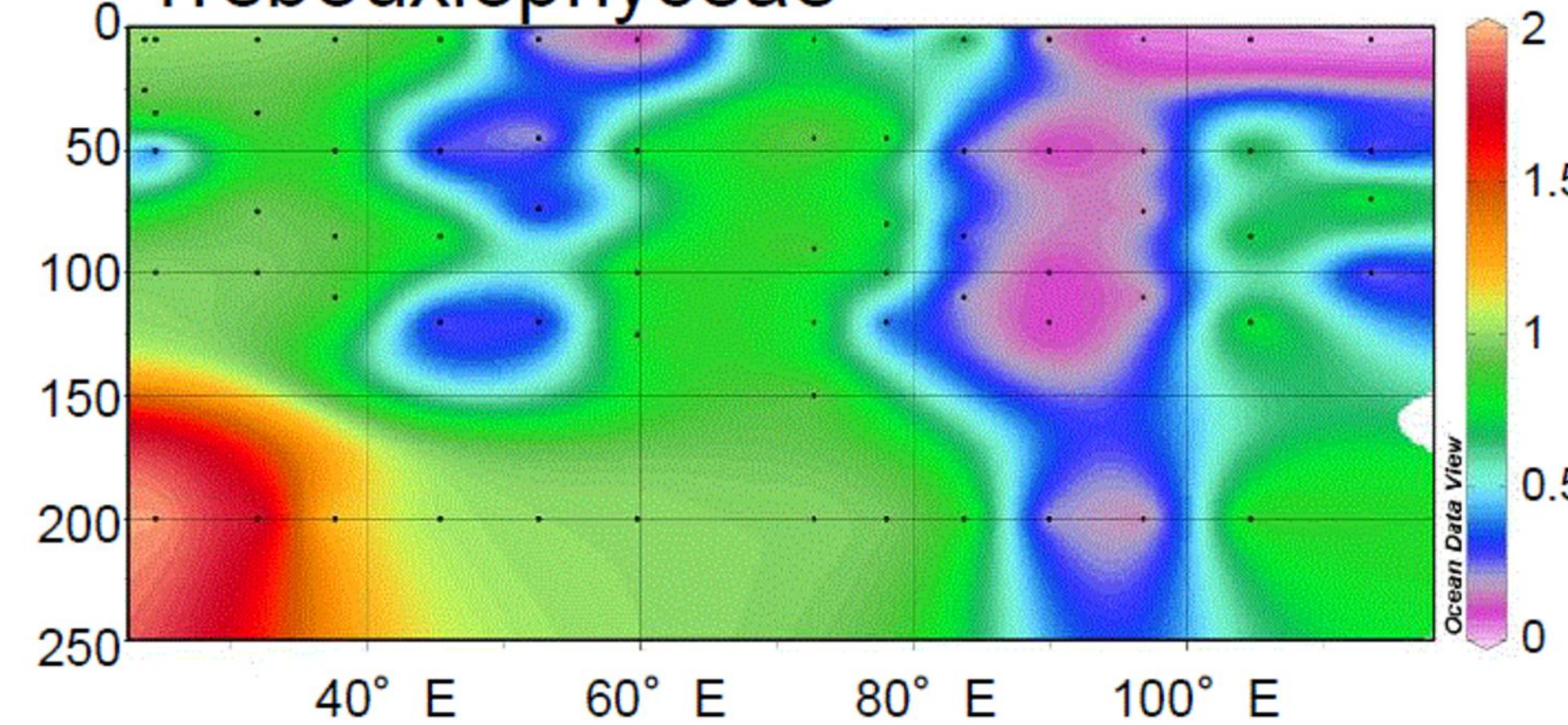
Pelagophyceae

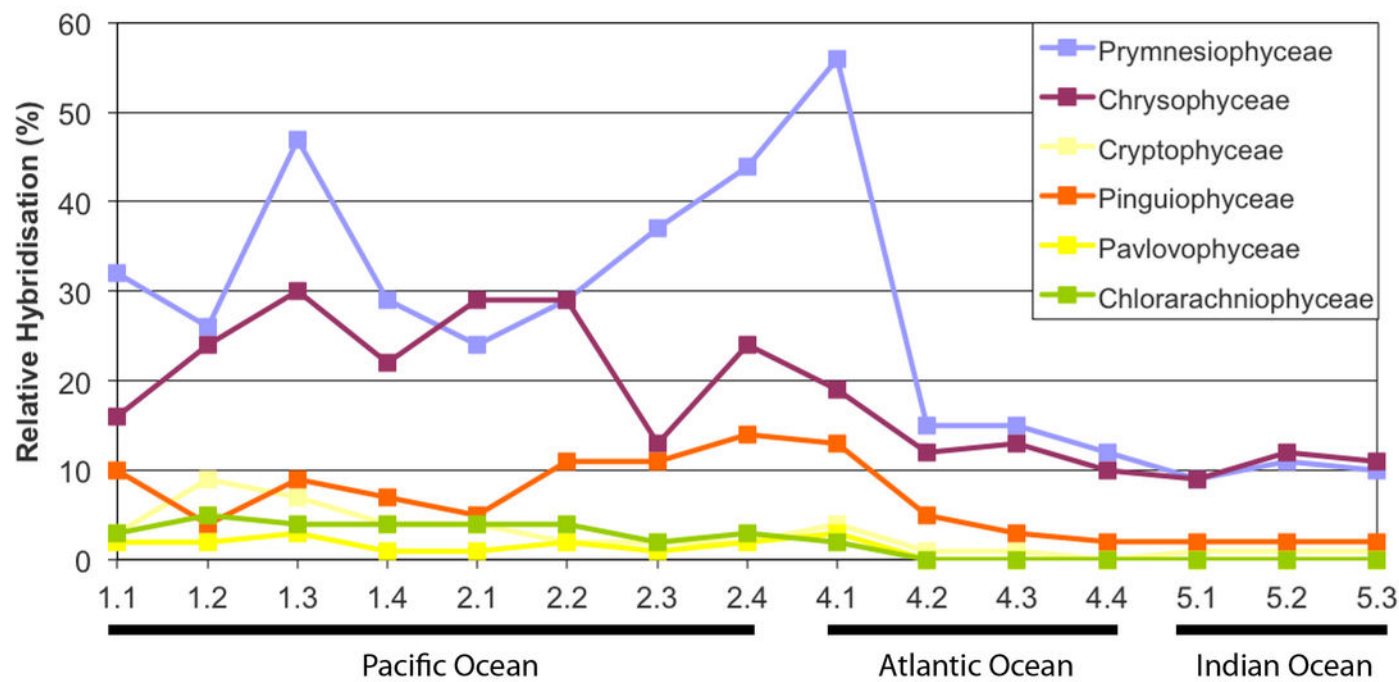
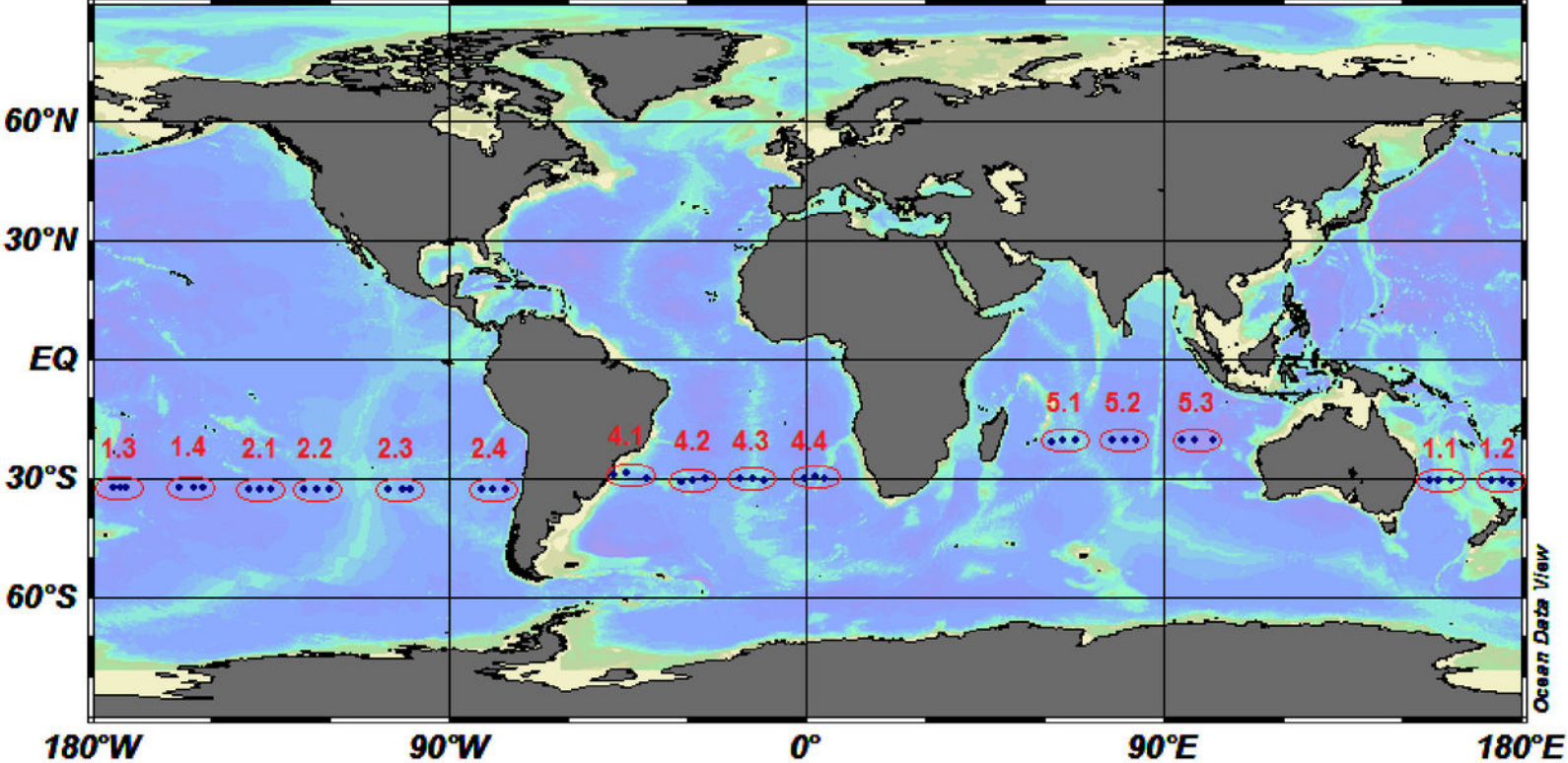


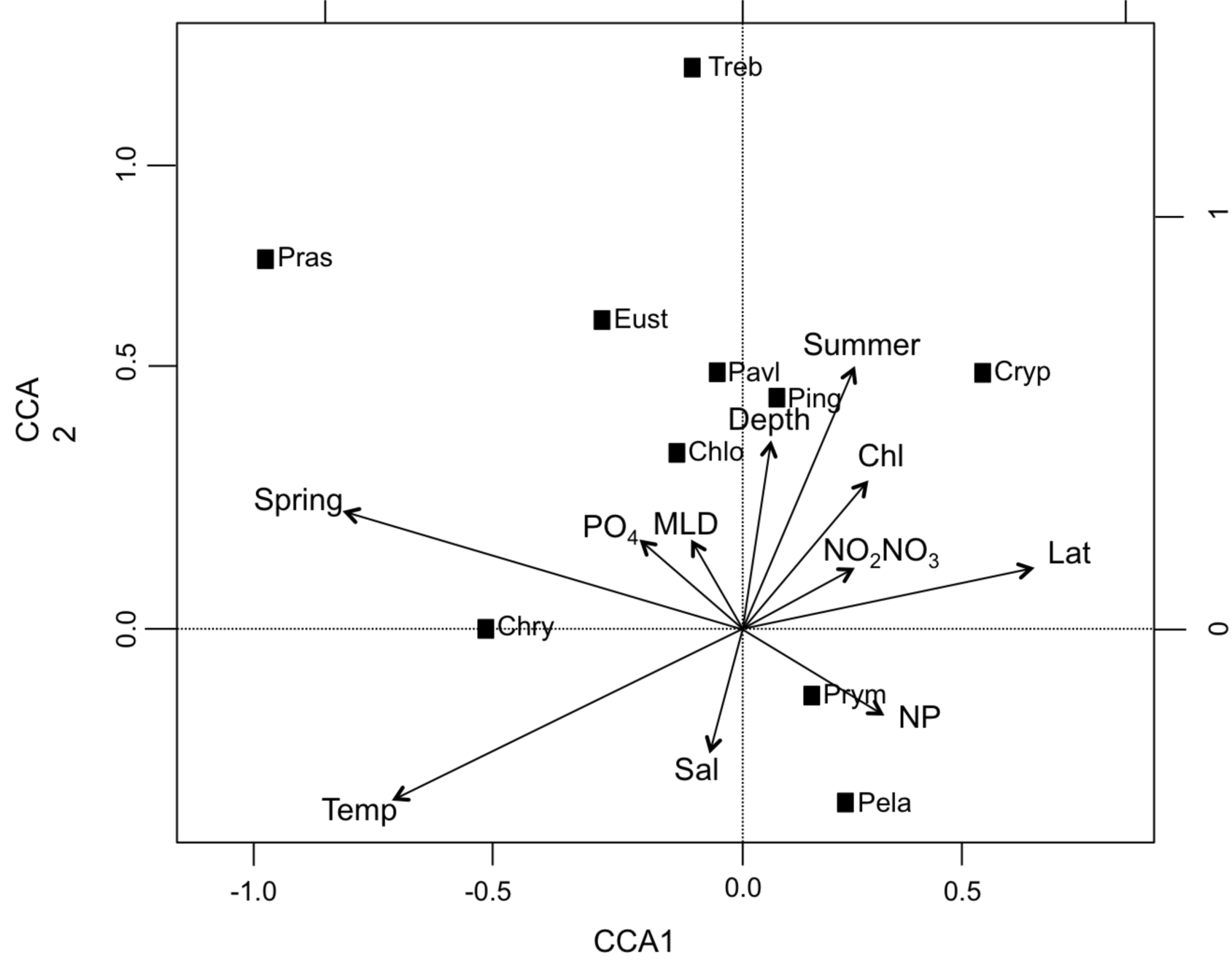
Pinguiphyceae



Trebouxiophyceae

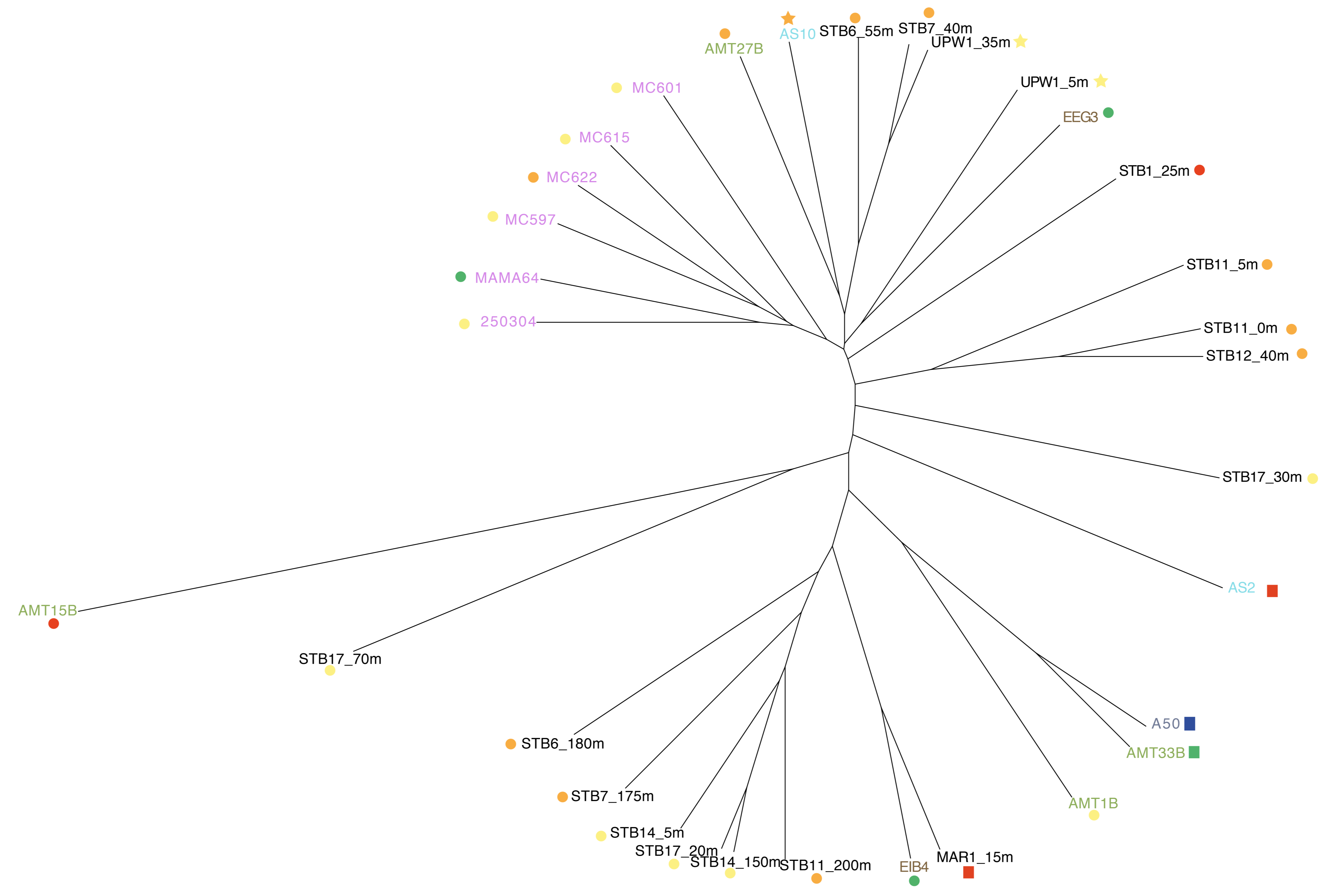


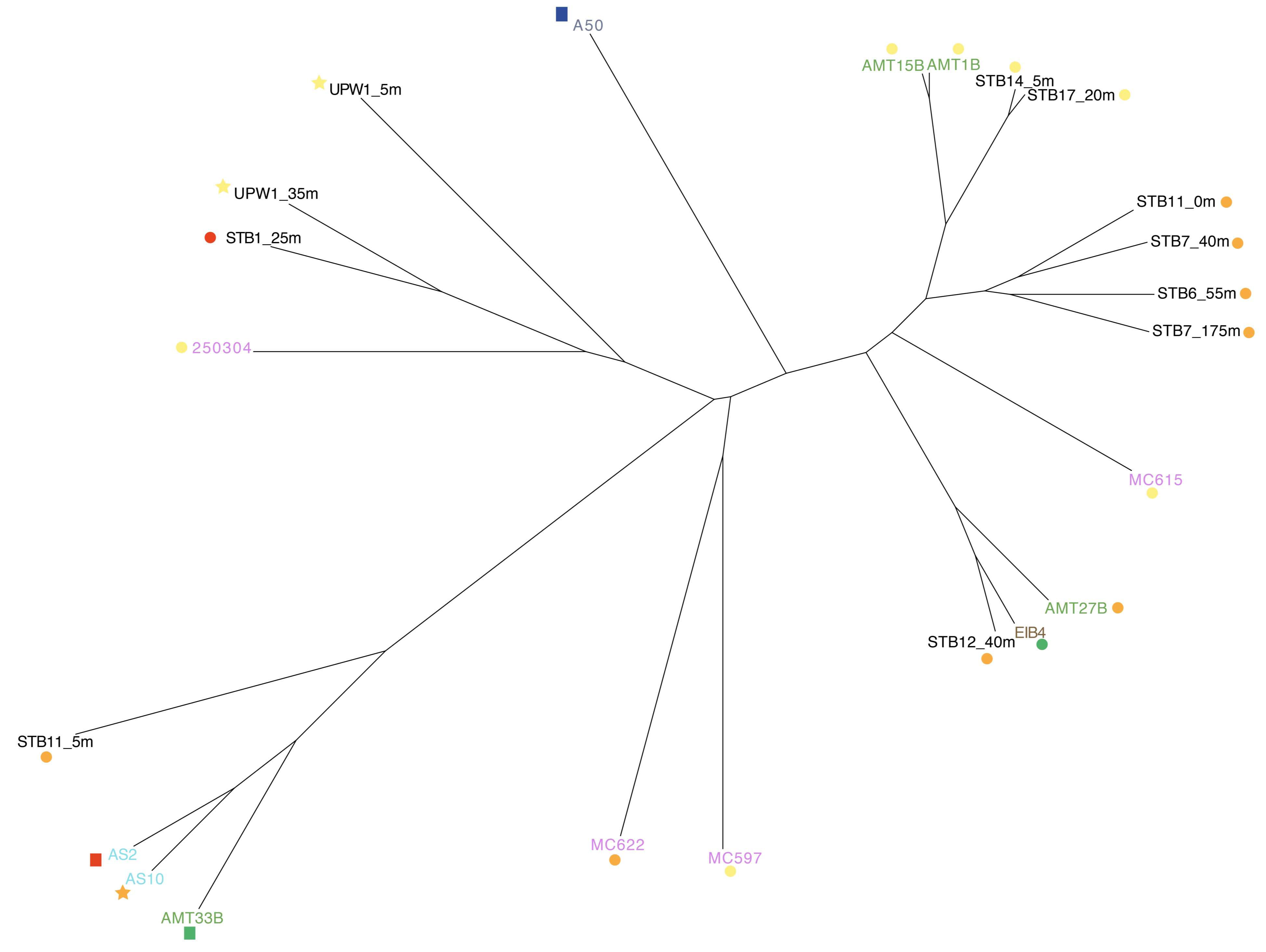












Cruise	Date	Geographic Region	Latitude/Longitude	Sampling	Reference
AMBITION	03/09/01 to 27/09/01	Arabian Sea	1°S-26°N, 56-67°E	11 stations 6 depths	Fuller <i>et al.</i> , 2006a
AMT15	19/09/04 to 26/10/04	Atlantic Meridional Transect	48°N-40°S, 25°W-10°E	21 stations 6 depths	Kirkham <i>et al.</i> , 2011a
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 work
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 work
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 work

**Table 1**

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

Table 2