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 β-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 3.9×10^4 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 8.2×10^2 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 9.2×10^2 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 multifactorial 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).

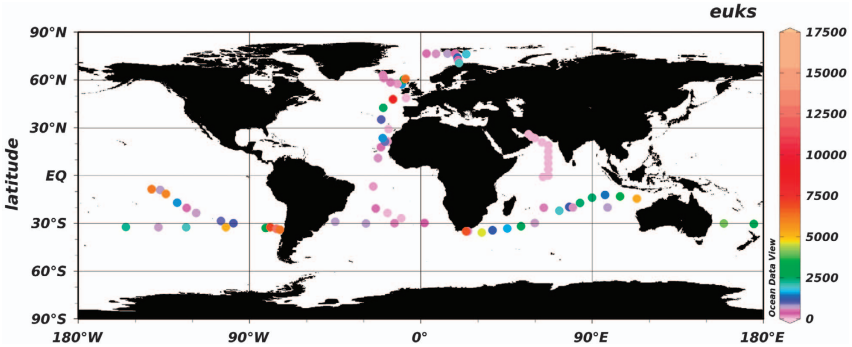

Figure 2 Global distribution patterns of PPE abundances, at all depths sampled, determined by flow cytometry.