environmental microbiology

Environmental Microbiology (2017) **19**(1), 287–300 doi:10.1111/1462-2920.13606

# Abundant and rare picoeukaryotic sub-communities present contrasting patterns in the epipelagic waters of marginal seas in the northwestern Pacific Ocean

**Wenxue Wu,[1,2,3] Ramiro Logares,[4] Bangqin Huang[1,2]* and Chih-hao Hsieh[3,5,6,7]***

[1]*State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China.*

[2]*Key Laboratory of Coastal and Wetland Ecosystems, Xiamen University, Xiamen, China.*

[3]*Institute of Oceanography, National Taiwan University, Taipei, Taiwan.*

[4]*Institut de Ciències del Mar (ICM), CSIC, Barcelona, Catalonia, Spain.*

[5]*Institute of Ecology and Evolutionary Biology, Department of Life Science, National Taiwan University, Taipei, Taiwan.*

[6]*Research Center for Environmental Changes, Taipei, Taiwan.*

[7]*National Center for Theoretical Sciences, Taipei, Taiwan.*

## Summary

**In this work, they compared patterns of abundant and rare picoeukaryotic sub-communities in the epipelagic waters (surface and 40–75 m depth subsurface layers) of the East and South China Seas across seasons via 454 pyrosequencing of the V4 region of 18S rDNA. They also examined the relative effects of environmental filtering, dispersal limitations and seasonality on community assembly. Their results indicated that (i) in the surface layer, abundant taxa are primarily influenced by dispersal limitations and rare taxa are primarily influenced by environmental filtering, whereas (ii) in the subsurface layer, both abundant and rare sub-communities are only weakly influenced by environmental filtering but are strongly influenced by dispersal limitations. Moreover, (iii) abundant taxa exhibit stronger temporal variability than rare taxa. They also found that abundant and rare sub-communities display similar spatial richness patterns that are negatively correlated with latitude and chlorophyll *a* and positively correlated with temperature. In summary, environmental filtering and dispersal limitations have different effects on abundant and rare picoeukaryotic sub-communities in different layers. Thus, depth appears as an essential variable that governs the structuring patterns of picoeukaryotic communities in the oceans and should be thoroughly considered to develop a more comprehensive understanding of oceanic microbial assemblages.**

Received 30 May, 2016; revised 2 November, 2016; accepted 8 November, 2016. *For correspondence: E-mail bqhuang@xmu.edu.cn (B. Huang): Tel, +86 592 2187783; Fax, +86 592 2180655 OR E-mail chsieh@ntu.edu.tw (C.-H. Hsieh): Tel, +886 2 33669745; Fax, +886 2 33669746.

© 2016 Society for Applied Microbiology and John Wiley & Sons Ltd

## Introduction

The majority of natural microbial communities seem to be composed of a few abundant taxa with ubiquitous distributions and a large number of rare species with restricted distributions (Pedrós-Alió, 2006; Logares *et al.*, 2015). Abundant microbes account for the majority of biomass and carbon cycling in ecosystems (Pedrós-Alió, 2012); thus, studying this community component is crucial to understand ecosystem functioning. However, recent studies have increasingly emphasized the importance of rare biota (Lynch and Neufeld, 2015), which include numerous metabolically active organisms (Campbell *et al.*, 2011; Debroas *et al.*, 2015). For instance, Pester *et al.* (2010) found that a rare bacterial species plays a crucial role in sulphate reduction in a peatland, and Musat *et al.* (2008) showed that a rare anaerobic phototrophic bacterial species is the primary facilitator of nitrogen and carbon cycling in Lake Cadagno. Moreover, even in highly diverse ecosystems, the rare biota supports critical ecological functions (Mouillot *et al.*, 2013); thus, the loss of rare species could have serious ecological consequences (Pendleton *et al.*, 2014).

Over the last decade, high-throughput sequencing has allowed researchers to analyse rare biota; thus, an increasing number of studies have compared abundant and rare taxa in various ecosystems (Logares *et al.*, 2014; Gong *et al.*, 2015; Liu *et al.*, 2015). Nevertheless, studies assessing the dynamics between abundant and rare marine picoeukaryotes ($<$ 3 μm in size) remain scarce.

Only one survey has been performed along European coastlines, and this study found that abundant and rare assemblages show contrasting patterns in terms of phylogenetic composition and community structure (Logares *et al*., 2014). This is curious because picoeukaryotes play a vital role in oceanic ecosystem functioning and are astoundingly diverse (Massana, 2011; de Vargas *et al*., 2015). This group of organisms includes parasitic groups (e.g., *Syndiniales* (*Dinophyta*, *Alveolata*), Guillou *et al*., 2008), photosynthetic producers, (e.g., *Mamiellophyceae* (*Chlorophyta*, *Archaeplastida*), Monier *et al*., 2016), mixotrophic taxa, (e.g., *Prymnesiophyceae* (*Haptophyta*, *Hacrobia*), Unrein *et al*., 2014), and heterotrophic organisms (e.g., MAST (*Stramenopiles*), Massana *et al*., 2014; *Radiolaria* (*Rhizaria*), Not *et al*., 2009). Thus far, whether abundant picoeukaryotes exhibit contrasting biogeographic patterns or different underlying community assembly mechanisms compared with rare picoeukaryotes in the open ocean remain unclear.

To understand the biogeography of marine picoeukaryotes, the mechanisms by which local biotic and abiotic conditions shape community structure must be understood, in particularly how they relate to environmental filtering. For instance, picoeukaryotes can serve as predators for bacteria and prey for larger phagotrophic protists (Caron *et al*., 1999), and they can exhibit distributions influenced by viral lysis (Mojica *et al*., 2016). Picoeukaryotes also exhibit spatial distributions that are closely associated with abiotic factors, such as nutrients (Painter *et al*., 2014). It is also necessary to consider that picoeukaryotes in epipelagic waters live in well-connected habitats, suggesting that dispersal among neighbouring communities can affect community structure (Hamilton *et al*., 2008). Such interconnections between communities in the sunlit ocean indicate that these groups can be treated as a metacommunity, which is a set of interacting communities that are linked by the dispersal of multiple, potentially interacting, taxa (Leibold *et al*., 2004). According to the metacommunity concept as well as an established hypothesis (Siqueira *et al*., 2012; Alahuhta *et al*., 2014), abundant picoeukaryotic sub-communities are more likely to be affected by local environmental conditions because their high abundances guarantee unrestricted dispersal; however, rare picoeukaryotic sub-communities are more likely to be affected by dispersal limitations (Siqueira *et al*., 2012).

Previous studies that have compared abundant versus rare picoeukaryotes in epipelagic waters did not consider the effects of ocean depth. Patterns of abundant versus rare picoeukaryotes in the surface layer have been documented in several regions (Kilias *et al*., 2014; Logares *et al*., 2014). However, the patterns of abundant versus rare picoeukaryotes in the subsurface layer (near the deep chlorophyll maximum [DCM]) are not well understood. We assume that abundant or rare picoeukaryotic sub-communities exhibit contrasting assembly mechanisms in the surface layer compared with the subsurface layers because the environmental conditions in these zones are distinctly different. The strength of environmental filtering and opportunities for dispersal most likely differ in the upper oceanic layers compared with deeper layers. For example, compared with the subsurface layer, the surface layer is commonly characterized by lower nutrient concentrations and stronger wind-induced mixing, which would result in greater environmental filtering and fewer dispersal limitations. Thus, it is relevant to compare the relative importance of environmental filtering and dispersal limitations in surface and subsurface layers for abundant or rare picoeukaryotes.

In this study, we investigated the diversity and distribution of picoeukaryotes using 50 samples collected from the surface (0–5 m depth) and subsurface (generally close to the DCM, ranging from 40 to 75 m depth) layers of epipelagic waters in the East China Sea (ECS) and South China Sea (SCS). Samples were collected across four different seasons, from 2009 to 2011 (Fig. 1). The key objectives of this study were to (1) evaluate how the interplay between environmental filtering and dispersal limitations influences the sub-community dynamics of abundant picoeukaryotes compared with rare picoeukaryotes; and (2) test whether abundant and rare picoeukaryotic diversity trends can be explained by selected environmental variables (e.g., temperature, latitude and chlorophyll *a*). We performed community pyrosequencing of the V4 region of 18S rDNA in multiple surface and subsurface samples. To our knowledge, the present research is the first to test this topic on a basin scale. In summary, this work expands our understanding of the forces driving community assembly in abundant or rare picoeukaryotic sub-communities in marine ecosystems.

## Results

### *Environmental characteristics*

The environmental conditions within the ocean's surface and subsurface layers varied significantly (Supporting Information Table S1). The surface layer was significantly warmer than the subsurface layer, which had average temperatures of 24.9°C and 22.9°C respectively ($P < 0.05$). Significantly lower salinities were observed in the surface layer (mean = 33.4) than in the subsurface layer (mean = 34.2) ($P < 0.05$). The differences in chlorophyll *a* were non-significant between the two layers ($P > 0.05$), whereas the nutrient levels ($NO_3 + NO_2$ and $PO_4$) were significantly higher ($P < 0.05$) in the subsurface layer (means = 4.1 and 0.3 μM respectively) than in the surface layer (means = 1.7 and 0.1 μM respectively). The mixed layer depths were relatively deeper during winter and fall (extending down to 83 and 92 m respectively) than in summer and spring

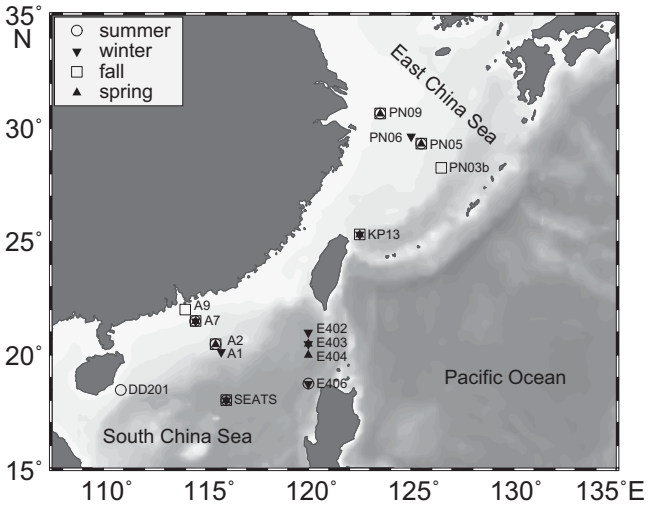


**Fig. 1.** Sampling locations in the East China Sea and the South China Sea from 2009 to 2011. Symbols indicate sampling cruises.

(extending down to 31 m and 18 m respectively), indicating stronger vertical mixing in winter and fall. In a principle component analysis (PCA), the first two axes (PC 1 and PC 2) used for subsequent variation partitioning accounted for the majority of the environmental variability identified (75.4% for surface waters; 70.5% for subsurface waters) (Supporting Information Fig. S1). These two axes were mainly associated to temperature, $NO_3 + NO_2$, $PO_4$ (PC 1), salinity and mixed layer depth (PC 2) in the surface layer (Supporting Information Fig. S1A), and temperature, $NO_3 + NO_2$, $PO_4$ (PC 1) and mixed layer depth (PC 2) in the subsurface layer (Supporting Information Fig. S1B).

### *Pyrosequencing characteristics*

Following our sequence-processing strategy, pyrosequencing yielded a total of 461497 high-quality sequences with an average length of approximately 420 bp, which were clustered into 4171 operational taxonomic units (OTUs; based on a 99% similarity threshold). Across all 50 samples, the top 20 OTUs had average relative abundances ranging from 0.7% to 8.3% and exhibited high similarity with sequences retrieved from GenBank (Supporting Information Table S2). The number of reads in each sample ranged from 4958 to 21231.

Our dataset was randomly subsampled 100 times to 4958 reads (the minimal read number across samples) per sample (Supporting Information Table S3) before downstream analyses. This approach (hereafter, bootstrap, *sensu* Yeh *et al.*, 2015) ensured that varying sequencing efforts per sample would not bias our results. From the 100 bootstraps, we identified abundant sub-communities composed of 150–165 OTUs (after removal of singletons) in the surface layer (from a total of 28 samples) and 109–125 OTUs in the subsurface layer (from a total of 22 samples). Correspondingly, we identified rare sub-communities composed of 3110–3178 OTUs in the surface layer and 3067–3143 OTUs in the subsurface layer.

### *Partitioning variance to examine community structure*

The results of a principal coordinate analysis (PCoA) based on unweighted UniFrac distances revealed that the majority of the differences between abundant versus rare sub-communities could be attributed to seasonality (Fig. 2). The first two PCoA axes explained more of the variances in the abundant sub-communities (55.6%, surface; 64.1%, subsurface; Fig. 2A and C) than in the rare sub-communities (23.8%, surface; 25.2%, subsurface; Fig. 2B and D).

The results of variation partitioning (3-way PERMANOVA) indicated that environmental, spatial and temporal factors contributed to the structure of abundant and rare sub-communities to different degrees (Table 1). In the surface layer, variation partitioning revealed that the exclusive environmental component [E|S + T] (i.e., the pure effects of environmental variables excluding spatial and temporal variables) possessed slightly more explanatory power in the rare sub-community (2.5%) than in the abundant sub-community (1.3%). In the subsurface layer, the exclusive environmental components presented limited contributions to either the abundant (0%) or rare (0.4%) sub-communities. The relative contribution of the exclusive spatial component [S|E + T] was slightly higher in the abundant (4.1%) than in the rare sub-communities (2.2%) in the surface layer, whereas in the subsurface layer, both abundant and rare sub-communities were significantly affected by this factor (10.5% and 7.3% respectively)

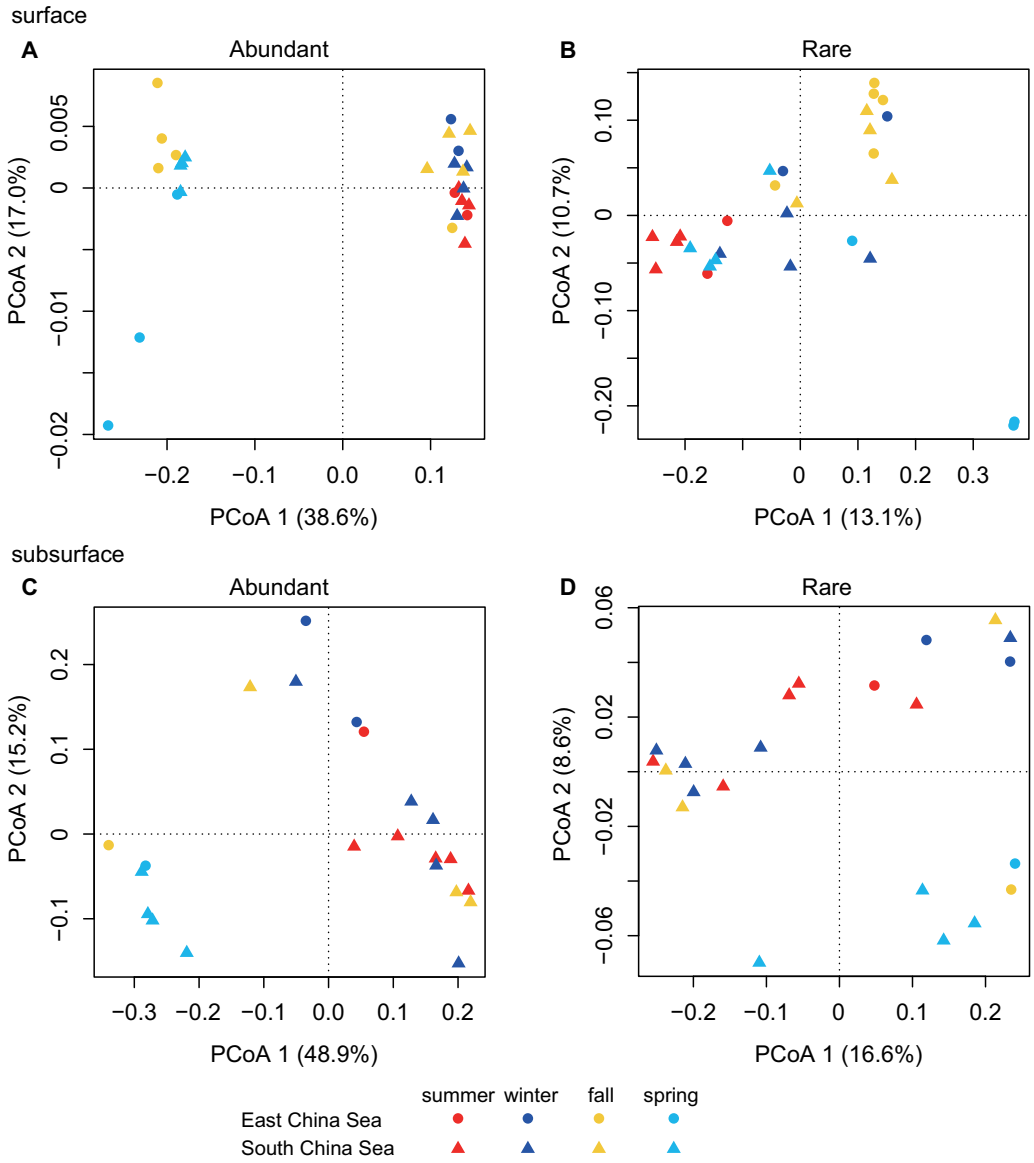


**Fig. 2.** Principal coordinate analysis (PCoA) of abundant and rare sub-communities in the surface (A and B) and subsurface (C and D) layers using an unweighted UniFrac distance metric. The results are based on the average of 100 bootstraps. [Colour figure can be viewed at wileyonlinelibrary.com]

($P < 0.05$ with 70% bootstrap support). Remarkably, both abundant and rare sub-communities were significantly affected by the exclusive temporal component [TIE + S] in the surface (32.7% for the abundant sub-community and 8.0% for the rare counterparts) and subsurface (41.8% for the abundant sub-community and 7.3% for the rare sub-community) layers.

### *Habitat specialization*

To help explain the differences in the structure of abundant or rare sub-communities, we examined their habitat specialization. We found that abundant taxa exhibited greater niche breadth index values (using each sample as an environmental type) in both the surface and subsurface layers (Fig. 3). In the surface layer, the average niche breadth for abundant sub-communities was 5.9 (SD = 0.5; $n = 28$) (Fig. 3A), whereas the average niche breadth for rare sub-communities was significantly lower (3.8; SD = 0.4; $P < 0.001$). In the subsurface layer (Fig. 3B), the average niche breadth for abundant sub-communities was 4.3 (SD = 0.6; $n = 22$), whereas for rare sub-communities, the average niche breadth was significantly lower (3.5; SD = 0.7; $P < 0.001$).

We further classified OTUs as generalists or specialists based on whether they presented a niche breadth which

**Table 1.** Variation partitioning results showing the contribution of environmental [E], spatial [S] and temporal [T] factors, as well as the unique environmental [E|S + T], unique spatial [S|E + T] and unique temporal [T|E + S] components of the abundant and rare sub-communities based on the unweighted UniFrac distance.

| | [E] | | [S] | | [T] | | [E\|S + T] | | [S\|E + T] | | [T\|E + S] | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | var | *P* | var | *P* | var | *P* | var | *P* | var | *P* | var | *P* |
| Surface | | | | | | | | | | | | |
| Abundant | **7.4** | **0.04** | 2.4 | 0.28 | **33.7** | **0.01** | 1.3 | 0.30 | 4.1 | 0.13 | **32.7** | **0.01** |
| Rare | **7.9** | **0.01** | **3.2** | **0.04** | **11.4** | **0.01** | 2.5 | 0.07 | 2.2 | 0.14 | **8.0** | **0.01** |
| Subsurface | | | | | | | | | | | | |
| Abundant | 0 | 0.51 | 7.8 | 0.09 | **32.3** | **0.01** | 0 | 0.49 | **10.5** | **0.03** | **41.8** | **0.01** |
| Rare | 2.0 | 0.14 | **6.4** | **0.01** | **5.3** | **0.02** | 0.4 | 0.52 | **7.3** | **0.01** | **7.3** | **0.01** |

*P*, the average of the100 bootstrapped *P*-value; var, the average of explained variance resulted from 100 bootstrapped 3-way PERMANOVA analyses. Bold values indicate strong statistical support for significant results ($P < 0.05$) with 70% bootstrap support.

was higher or lower than chance respectively. On average, abundant sub-communities were composed of 17.5% generalists and 21% specialists in the surface layer and 21.9% generalists and 20.1% specialists in the subsurface layer (Supporting Information Fig. S2). Rare sub-communities included 4.6% generalists and 4.9% specialists in the surface layer and 3.9% generalists and 4.3% specialists in the subsurface layer.

### *Community composition*

To further understand abundant or rare sub-community dynamics, we also investigated community taxonomic composition, because life history strategies vary among taxonomic groups and are often associated with habitat utilization (Marzluff and Dial, 1991). By analysing the relative OTU distributions, we found that abundant sub-communities in the surface layer were generally composed of 68.3% *Dinophyta*, 11.1% *Chlorophyta* and 6.9% *Haptophyta* OTUs (Fig. 4A). *Dinophyta* also accounted for the largest fraction of OTUs (56.3%) in the 50 rare sub-communities identified in the surface layer, and it was followed by *Stramenopiles* and *Ciliophora* at an average of 10.3% and 8.6% respectively. In the subsurface layer (Fig. 4B), *Radiolaria* accounted for 17.3% and 11.5% of the OTUs in abundant and rare sub-communities respectively.

In terms of group abundances (using number of reads as a proxy of organismal abundances) *Dinophyta* dominated in both the surface (66.9% for abundant and 56.5% for rare sub-communities) and subsurface (43.1% for abundant and 62.9% for rare sub-communities) layers (Supporting Information Fig. S3). *Radiolaria* were relatively more abundant in the subsurface layer (27.2% and 14% of abundant and rare sub-communities respectively) (Supporting Information Fig. S3B and D) than in the surface layer (3.1% and 4.6% of abundant and rare sub-communities respectively) (Supporting Information Fig. S3A and B). Notably, *Chlorophyta* were present at significantly greater levels in the abundant sub-communities (19.9% in surface and 23.8% in subsurface communities,

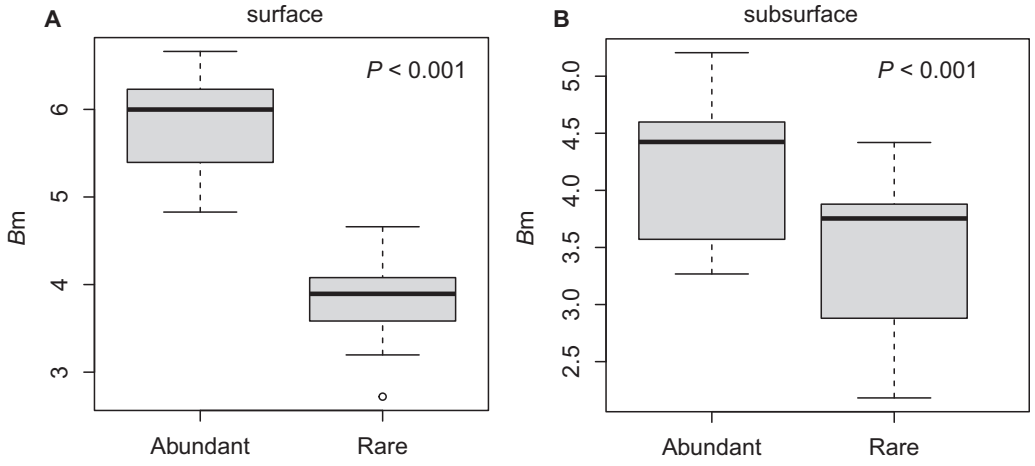


**Fig. 3.** Boxplots summarizing mean niche breadth (*B*m) for each abundant and rare sub-community sample in the surface layer ($n = 28$; $P < 0.001$) (A) and subsurface layer ($n = 22$; $P < 0.001$) (B). The *B*m value was calculated using the mean niche breadth (*B*) of all taxa in a sub-community. The results are based on the average of 100 bootstraps.

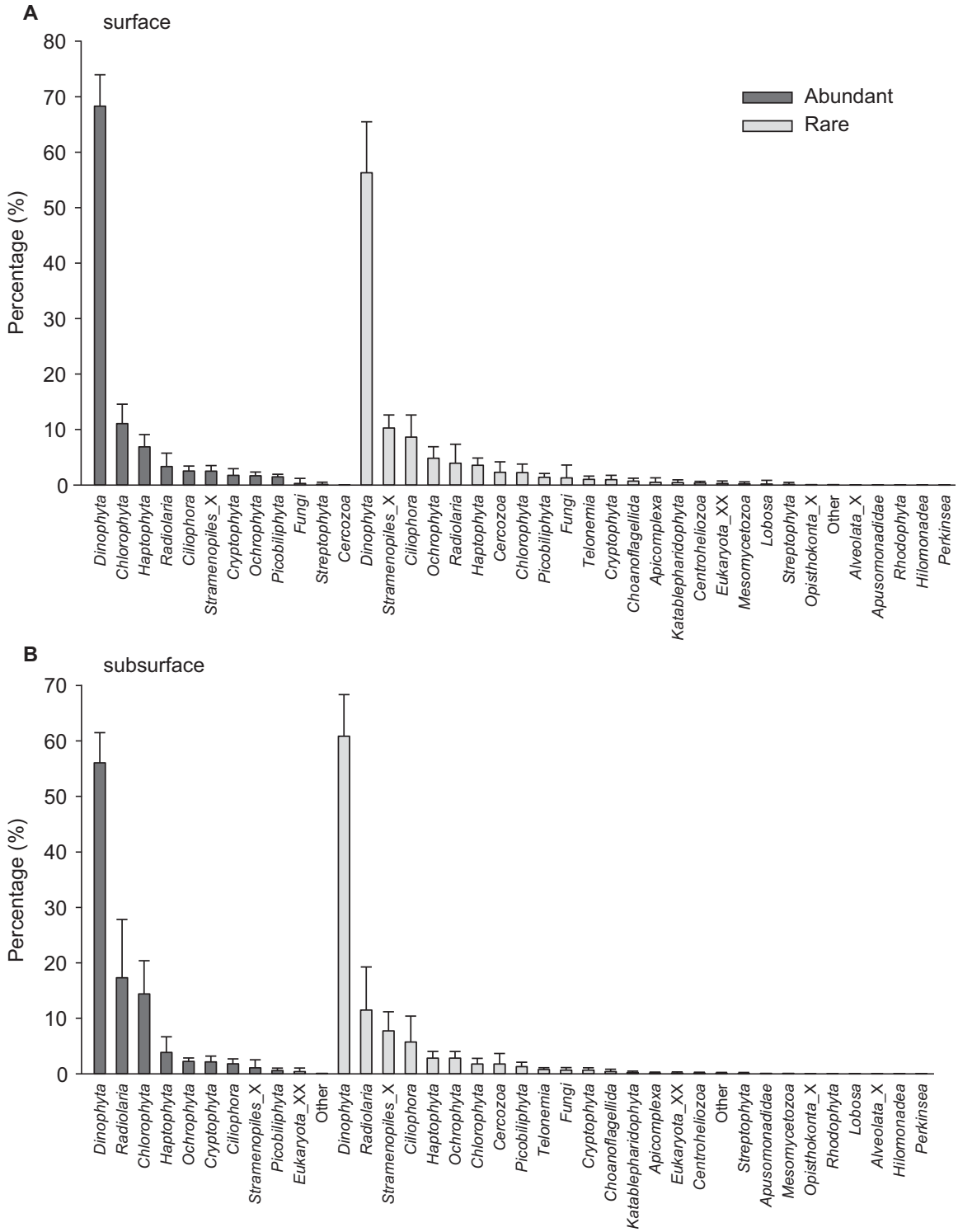


**Fig. 4.** Relative contributions of the taxonomic groups in the abundant and rare sub-communities in the surface (A) and subsurface (B) layers. Each bar shows the average value across the surface samples ($n = 28$) and subsurface samples ($n = 22$) with a standard deviation. The results are based on the average of 100 bootstraps.

Supporting Information Fig. S3A and B) relative to rare sub-communities (2.7% in surface and 1.5% in subsurface communities, Supporting Information Fig. S3C and D).

### *Phylogenetic diversity in relation to environmental variables*

Along with the phylogenetic metric for beta diversity (unweighted UniFrac), we analysed the phylogenetic diversity (PD) (Faith, 1992), which can be considered the phylogenetic counterpart to species richness. PD sums the branch lengths within each sample (Tucker *et al.*, 2016). The PD index for abundant versus rare sub-communities were correlated with latitude, temperature and chlorophyll *a* concentrations (Fig. 4). In the surface layer, the PD of abundant taxa exhibited a significantly positive relationship with temperature ($r = 0.52$, $P < 0.05$ and bootstrap support value = 100) and a significantly negative relationship with chlorophyll *a* concentrations ($r = -0.45$, $P < 0.05$ with a bootstrap support value of 100). For rare taxa, the PD index showed slightly larger significant correlations with temperature ($r = 0.59$, $P < 0.05$ and bootstrap support value = 100) and chlorophyll *a* concentrations ($r = -0.51$, $P < 0.05$ and bootstrap support value = 100) than that of the abundant taxa. In the subsurface layer, similar but non-significant correlations (less than 70 out of 100 bootstraps with $P < 0.05$) were observed between the PD indices of abundant and rare taxa versus temperature ($r = 0.22$, 0 bootstrap with $P < 0.05$ for abundant taxa; $r = 0.22$, 0 bootstrap with $P < 0.05$ for rare taxa) and chlorophyll *a* ($r = -0.51$, 0 bootstrap with $P < 0.05$ for abundant taxa; $r = -0.13$, 0 bootstrap with $P < 0.05$ for rare taxa). At both depths, the PD of abundant and rare sub-communities exhibited a non-significant decline with increasing latitude.

## Discussion

### *Determinants of abundant and rare sub-community dynamics*

The contrasting sub-community dynamics of abundant versus rare picoeukaryotes were primarily related to environmental filtering or dispersal limitations. Unexpectedly, the results from the surface layer analyses did not support our initial hypothesis of community assembly but rather indicated that abundant sub-communities in this layer are influenced more by dispersal limitations than rare sub-communities and that rare sub-communities are more responsive to environmental filtering than their abundant counterparts (Table 1). Nevertheless, these results are consistent with a recent study of inland lakes and reservoirs, which demonstrated that abundant bacterial sub-communities are primarily driven by regional dispersal factors, whereas rare bacterial sub-communities are primarily governed by local environmental variables (Liu *et al.*, 2015). Our data indicate that abundant picoeukaryotes have wider niche breadths than rare picoeukaryotes (Fig. 3) and show that abundant sub-communities contain a higher proportion of habitat generalists than their rare counterparts (Supporting Information Fig. S2). These results (as well as those in Liu *et al.*, 2015) may be explained by an alternative hypothesis proposed by Pandit *et al.* (2009), which suggests that taxa with a wide niche breadth are primarily influenced by dispersal limitations, whereas the distribution of taxa with a narrow niche breadth are more affected by environmental filtering. This alternative hypothesis argues that the habitat occupancy of taxa with wide niche breadth may be limited by the possibilities to reach multiple locations (dispersal limitation). On the other hand, taxa with narrow niche breadths may face strong negative environmental selection.

This alternative hypothesis may be particularly relevant when connectivity among patches is high. Indeed, marine patches in the surface layer are often well connected. The East China Sea (ECS) and South China Sea (SCS) are characterized by active surface currents forced by the East Asia monsoonal winds (Hu *et al.*, 2000; Ichikawa and Beardsley, 2002). Interestingly, our results suggest that the strength of dispersal limitations may be different for abundant and rare taxa. Taxa with high local abundances at one site appear to be present at a number of adjacent sites as a result of prevalent current transportation, whereas taxa with extremely low abundances are less likely to move from one site to another. Such a positive abundance-occurrence relationship has been observed for bacteria and phytoplankton (Östman *et al.*, 2010) and may result in stronger responses to dispersal limitations in abundant picoeukaryotic sub-communities than in rare sub-communities. Even though the ECS and SCS are connected, they have separate surface currents that may serve as a barrier and limit dispersal. This could explain the greater dissimilarities between abundant sub-communities when compared with rare counterparts (Fig. 2A and B). Overall, the differences in niche breadths between abundant and rare picoeukaryotes may explain the variable contributions of environmental filtering and dispersal limitations to the sub-community dynamics in the surface layer, where patch connectivity is high because of wind-driven surface currents.

However, our subsurface layer results do not support either of the aforementioned hypotheses. Instead, our findings showed that abundant and rare sub-communities are similarly affected by exclusive environmental and spatial components (Table 1). Importantly, the effects of dispersal limitations are stronger in the subsurface layer than in the surface layer, which may be explained by the weak horizontal mixing in the subsurface layer, which heavily limits dispersal (Hanson *et al.*, 2012). Moreover, the subsurface

layer in this study is close to the DCM, which is characterized as a hotspot of biotic activity (Venrick *et al*., 1973). Such complex biotic activities were not measured in this study. Therefore, the apparent lack of the influence of environmental filtering in the subsurface layer may be related to unconsidered biotic interactions.

The structures of both abundant and rare sub-communities were partially impacted by temporal factors, although this impact is more evident in abundant taxa (Table 1). Furthermore, because both abundant and rare picoeukaryotes are metabolically active (Campbell *et al*., 2011; Logares *et al*., 2014), the community dynamics in both picoeukaryote groups can be strongly influenced by temporal factors. In the present study, these influences were observed in the community responses to seasonality in the ECS and SCS. Indeed, seasonal occurrences, such as those associated with atmospheric inputs are well known in the ECS and SCS (Gao *et al*., 1992; Kim *et al*., 2011). Our results are consistent with previous research showing that both abundant and rare bacteria can exhibit significant temporal patterns (Alonso-Sáez *et al*., 2015). In our study, rare species are defined as those that occur infrequently during all four seasons, which reflects a low ability to compete (Hugoni *et al*., 2013), population control by predators or virus/parasites, or intrinsically low growth rates (Logares *et al*., 2015). It should be noted that rare species may be stochastically abundant in terms of dormancy and 'boom and bust' episodes (Alonso-Sáez *et al*., 2015), although our limited sample sets, which were gathered during four discrete cruises, are not sufficient to determine whether such factors are relevant here. However, abundant and competitive superior species tend to be responsive to environmental variability. Thus, we suggest that the differences in competition for resources between abundant and rare picoeukaryotic sub-communities and the temporal environmental variability within the ECS and SCS result in greater temporal influences on abundant sub-community assembly than on their rare counterparts.

In this study, we evaluated the relative effects of environmental filtering, dispersal limitations and seasonality on the assembly of abundant and rare picoeukaryotic sub-communities using variation partitioning. However, we acknowledge that there were limitations to our sampling method; for example, we did not collect all samples at the same approximate time, and we did not sample on a fixed regular grid. As a consequence, our samples may not wholly reflect the influences of environmental variability and dispersal processes on community assembly. Nevertheless, this research represents the first large-scale cross-basin survey in the northwest Pacific Ocean, and the sampled stations reflect wide spatial distributions across all seasons.

### *Compositional variability of abundant versus rare sub-communities*

Abundant and rare picoeukaryotes display contrasting compositional patterns, which provides information on their different sub-community compositions. In the surface layer, abundant sub-communities contain more photosynthetic groups, whereas rare sub-communities are consistently and primarily composed of heterotrophic groups (Fig. 3). The majority of the photosynthetic abundant taxa fall within three *Mamiellophyceae* (*Chlorophyta*, *Archaeplastida*) genera: *Ostreococcus*, *Micromonas* and *Bathycoccus* (Supporting Information Table S1). *Ostreococcus*, *Micromonas* and *Bathycoccus* are composed of a number of phylotypes that are widespread in the oceans (Monier *et al*., 2016), suggesting wide niche breadths for the members of these groups. When niche breadth and local abundance are taken into account (Supporting Information Fig. S3), these three genera are likely to account for most of the observed dispersal limitation patterns. Hence, we suggest that patterns pointing to higher dispersal limitations in the surface layer may be closely related to the ecology of the predominant *Ostreococcus*, *Micromonas* and *Bathycoccus*.

In the subsurface layer, the abundant and rare sub-communities were composed of relatively greater levels of *Radiolaria* than the same communities in the surface layer (Fig. 4). This could explain to certain extent the observed patterns of dispersal limitations, as subsurface waters provide less opportunities for horizontal dispersal than surface waters. Many *Radiolaria* species (including *Acantharia* and *Polycystinea* species) have the capacity to secrete $SrSO_4$ to build skeletons and individual crystals (Martin *et al*., 2010). As such, cyst-forming radiolarian cells (e.g., *Acantharia* Clades A, B and C) may rapidly sink from the subsurface to deeper layers within a life cycle (Decelle *et al*., 2013), thus limiting horizontal dispersal within a single layer. The majority of *Radiolaria* sequences in this study belonged to cyst-forming *Spumellarida* (*Polycystinea*) (Supporting Information Table S1), which have a life cycle similar to that of cyst-forming *Acantharia* (Yuasa and Takahashi, 2016).

### *PD dynamics across environmental gradients*

Our findings show that the PD of both abundant and rare taxa is positively correlated with temperature (Fig. 5B and E) and negatively correlated with chlorophyll *a* (Fig. 5C and F) in the surface layer, indicating that the diversity patterns of both abundant and rare picoeukaryotes may be related to temperature and chlorophyll *a*. However, the abundant picoeukaryotes responded more weakly to temperature and chlorophyll *a* concentration gradients than the rare picoeukaryotes. The latter could have been

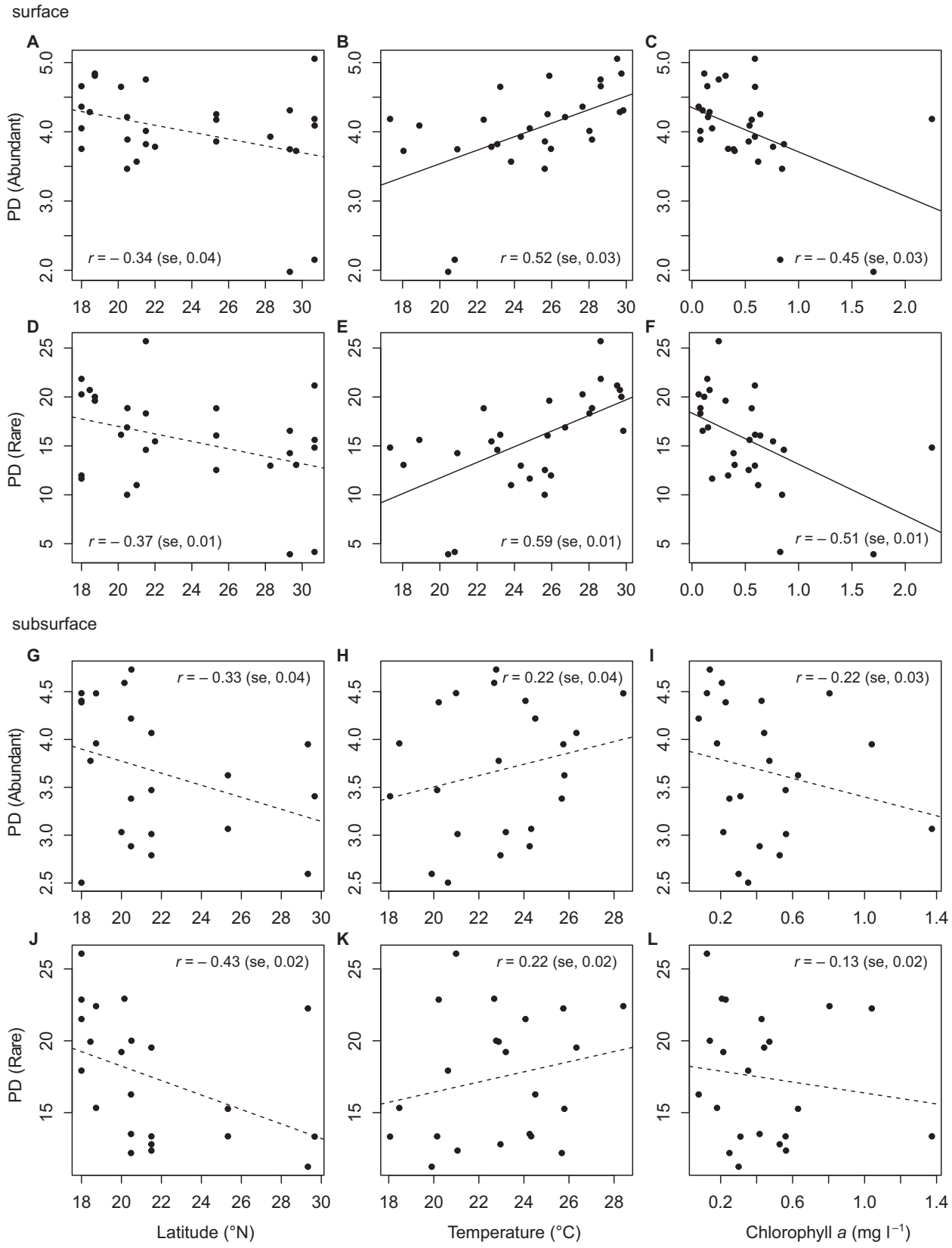
surface
A
B
C
D
E
F
PD (Abundant)
PD (Rare)
r = – 0.34 (se, 0.04)
r = 0.52 (se, 0.03)
r = – 0.45 (se, 0.03)
r = – 0.37 (se, 0.01)
r = 0.59 (se, 0.01)
r = – 0.51 (se, 0.01)
subsurface
G
H
I
J
K
L
r = – 0.33 (se, 0.04)
r = 0.22 (se, 0.04)
r = – 0.22 (se, 0.03)
r = – 0.43 (se, 0.02)
r = 0.22 (se, 0.02)
r = – 0.13 (se, 0.02)
Latitude (°N)
Temperature (°C)
Chlorophyll a (mg l−1)

**Fig. 5.** Correlation between Faith's phylogenetic diversity (PD) index for abundant and rare picoeukaryotic sub-communities and latitude (A, D, G and J), temperature (B, E, H and K), and chlorophyll *a* concentrations (C, F, I and L). Dots represent the average of 100 bootstrapped PD metrics. The statistic value of the Pearson test ($r$) and the standard error (se) are presented. Solid (dash) regression lines indicate strong (weak) statistical support of significant results ($P < 0.05$) for more (less) than 70 out of 100 bootstraps.

generated by a decrease of beta diversity produced by high dispersal levels in abundant picoeukaryotes (Hubert *et al.*, 2015). In the subsurface layer (Fig. 5G–L), the correlations between PD and latitude, temperature and chlorophyll *a* were non-significant. We suggest that this lack of significance may result from complex biotic activities such as prey-predator interactions (beyond latitude, temperature and chlorophyll *a*) around the subsurface DCM that limits the environmental effects.

## Conclusions

Variable environmental filtering and dispersal limitations in the surface and subsurface layers of the ECS and SCS generate different sub-communities of abundant or rare picoeukaryotes. In the surface layer, where connectivity is high because of strong surface currents, the difference in niche breadth (e.g., a wide niche breadth for abundant picoeukaryotes and a narrow niche breadth for rare picoeukaryotes) may explain the variation in the influence of environmental filtering and dispersal limitations on the picoeukaryotic sub-community dynamics. In the subsurface layer, both abundant and rare sub-communities were greatly impacted by dispersal limitations but only weakly affected by environmental filtering, which may be related to weak horizontal mixing and complex biotic activities. The richness patterns for both abundant and rare picoeukaryotes were negatively correlated with latitude and chlorophyll *a* and positively correlated with temperature, although at different magnitudes. These findings contribute to the scientific community's understanding of picoeukaryotic community assembly in the epipelagic waters of marine ecosystems.

## Experimental procedures

### *Sample collection and sequencing*

Sampling was conducted in the East China Sea (ECS) and South China Sea (SCS) during four different seasonal cruises from 2009–2011 (Fig. 1). The ECS and SCS are two large marginal seas located in the western Pacific Ocean extending from 23°N to 33°N (117–131°E) and 3°S to 23°N (99–121°E) respectively. A total of 50 samples were collected from the surface and subsurface (63 m average, close to the DCM) layers of the epipelagic waters of both seas (Supporting Information Table S1). A majority of samples were collected using Niskin bottles mounted on a CTD (conductivity, temperature and depth) rosette (Sea-Bird Electronics, USA), and several surface samples were collected at a depth of 0 m with a clean bucket (Supporting Information Table S1). Briefly, seawater samples (2–10 l) from the fall, winter, and spring cruises were pre-filtered through 3-μm pore size polycarbonate membranes (Millipore, USA) and from the summer cruise were pre-filtered through 5-μm pore size filters. The pre-filtered seawater samples were subsequently filtered through GF/F filters (Whatman, USA) with 200-mm Hg. The filters were immediately stored at −80°C until further analysis.

Total DNA was extracted following the phenol:chloroform:isoamylalcohol method (Countway *et al.*, 2005). PCR analyses were performed using 454 sequencing adaptors linked with primers for the V4 hypervariable region of the eukaryotic SSU rDNA (TAReuk454FWD1 5′-CCAGCA(G/C)C(C/T)GCGG TAATTCC-3′ and TAReukREV3 5′-ACTTTCGTTCTTGAT(C/-T)(A/G)A-3′) (Stoeck *et al.*, 2010). For each sample, triplicate amplifications were conducted in 20-μl reactions using TransStart Fastpfu DNA polymerase (TransGen, China) on a G eneAmp 9700 system (ABI, USA). The PCR program consisted of an initial denaturation step at 95°C for 2 min; 10 cycles at 95°C for 30 s, 53°C for 30 s and 72°C for 1 min; and then 20 additional cycles at 95°C for 30 s, 48°C for 30 s and 72°C for 1 min. A final extension occurred at 72°C for 5 min. The PCR products were purified from a 2% agarose gel using an AxyPrep DNA Gel Extraction Kit (Axygen, USA). Pyrosequencing was performed using a 454 GS FLX Titanium system (Roche, USA) following the manufacturer's instructions. The raw pyrosequencing sequences have been deposited in the sequence read archive of GenBank under the accession numbers SRP070982 and SRP071043.

### *Sequence processing*

Sequence processing was performed using the Quantitative Insights Into Microbial Ecology (QIIME v. 1.8.0) pipeline (Caporaso *et al.*, 2010). First, the quality of reads was checked using a 50-bp sliding window, and reads with an average Phred score $< 25$ were discarded. Sequences between 200 and 500 bp were used for the downstream analyses. All of the pyrotags were then denoised using the DeNoiser (Reeder and Knight, 2010) integrated in QIIME. Quality trimmed reads were clustered into OTUs using a 99% similarity threshold with UCLUST (Edgar, 2010). Chimeras were identified using ChimeraSlayer (Haas *et al.*, 2011). Subsequently, representative reads for the phylotype OTUs were assigned using BLAST (Zhang *et al.*, 2000) within QIIME ($E$ value $= 10^{-6}$; minimum percent similarity $= 90\%$) against the $PR^2$ database based on GenBank 203 (Guillou *et al.*, 2013). OTUs representing singletons were removed from the OTU table along with those present in a single sample. The OTUs assigned to metazoan were also excluded. To obtain a clean dataset, a phylogenetic tree was constructed in QIIME using the representative sequences, and each OTU in a long branch was manually checked using BLAST.

### *Environmental variables*

The hydrography of each station was determined using a CTD cast. The micronutrient concentrations (nitrate + nitrite, phosphate) were measured using a Technicon AA3 Auto-Analyser (Bran-Lube, Germany) as described in detail in Du *et al.* (2013). The chlorophyll *a* concentrations were determined using an acetone extraction method (Wu *et al.*, 2014) and high-performance liquid chromatography (Wu *et al.*, 2015). The mixed layer depth was calculated as the first depth at which the temperature was 0.2°C lower than the temperature at the surface for that station (Steinhoff *et al.*, 2010).

### *Defining abundant versus rare taxa*

In this study, OTUs were considered abundant when they accounted for $> 1\%$ of the relative contribution within any sample, and they were considered rare if they accounted for $< 1\%$ in all samples (Campbell *et al.*, 2011; Vergin *et al.*, 2013; Alonso-Sáez *et al.*, 2015). Abundant and rare sub-communities were separately defined for the surface and subsurface layers. When defining the abundant and rare OTUs, each sample was randomly subsampled to the minimum read number of all samples to avoid biases generated by differences in sequencing depth (Yeh *et al.*, 2015). We randomly repeated the subsampling 100 times (Supporting Information Appendix S1) to ensure that the subsampling procedure would not bias the downstream analyses based on relative abundance. Thus, all of the subsequent analyses, which relied on a subsampled OTU table to address abundant and rare taxa, were repeated 100 times.

### *Variation partitioning of the environmental, spatial and temporal components*

We used the variation partitioning approach (Borcard *et al.*, 1992) to determine the relative effects of the environmental, spatial and temporal factors on the community variations. To perform variation partitioning, a 3-way PERMANOVA (McArdle and Anderson, 2001) was applied to decompose the total variance of the community differentiation into relevant components. Phylogenetic-based beta diversity was calculated using the UniFrac algorithm (visualized using PCoA plots) to analyse the variability in the phylogenetic composition of communities across multiple ecological gradients (Lozupone and Knight, 2005). Specifically, a phylogenetic tree was constructed using FastTree (Price *et al.*, 2009). To construct the tree, representative OTU sequences were aligned against the SILVA 111 database (Quast *et al.*, 2013) using mothur (Schloss *et al.*, 2009). The pairwise community dissimilarity among sites was then measured using the UniFrac distance metric. In this study, unweighted UniFrac distances were used because metacommunity analyses commonly focus on the presence-absence of taxa (e.g., Beisner *et al.*, 2006; Yeh *et al.*, 2015). The unweighted UniFrac-based community dissimilarities were visualized using a principal coordinate analysis.

To determine the underlying mechanisms that shape community assembly, environmental heterogeneity and dispersal limitations were considered (Leibold *et al.*, 2004; Cottenie, 2005). To determine environmental heterogeneity, a PCA was performed using the R package vegan v. 2.2 (Oksanen *et al.*, 2014) based on environmental variables, including the physical, chemical and biological parameters mentioned above. The first two axes were then selected according to the Kaiser–Guttman rule. For the dispersal limitations, we relied on Moran's eigenvector maps (MEMs) (Dray *et al.*, 2006) based on the latitude and longitude coordinates of our sampling locations and used the R package PCNM v. 2.1-2 (Legendre and Gauthier, 2014). To define the temporal variability, MEMs were also constructed according to the sampling season. Because of the seasonal events that occur in the ECS and SCS (e.g., atmospheric depositions, Gao *et al.*, 1992; Kim *et al.*, 2011; surface currents, Hu *et al.*, 2000; Ichikawa and Beardsley, 2002), time was suggested to represent the unmeasured environmental components (such as other environmental predictors not collected in this study) and dispersal limitations (such as current transport, which was overlooked because of the limited sampling effort). Using unweighted UniFrac dissimilarities, a 3-way PERMANOVA was performed with an ad hoc R script following Yeh *et al.* (2015) (Supporting Information Appendix S2), and the bootstrap support for variation partitioning (out of 100) was reported. The statistical analyses were conducted using the R program (R Core Team, 2014).

### *Habitat specialization*

To help explain the patterns of beta diversity, we estimated Levins' niche breadth (*B*) index for the abundant and rare subcommunities (Levins, 1968):

$$B_j = \frac{1}{\sum_{i=1}^{N} P_{ij}^2}$$

where $B_j$ indicates the niche breadth of OTU *j*; *N* is the total number of communities; and $P_{ij}$ is the proportion of OTU *j* in community *i*. The calculations of $B_j$ for the surface and subsurface samples were performed separately ($N = 28$ and 22 in the surface and subsurface samples respectively). The *B* index considered habitat utilization based on the species abundance and evenness at a metacommunity scale. Note, a higher *B*-value indicates a wider niche breadth for species that should occur in more habitats; a lower *B*-value indicates a narrower niche breadth for those species that occur in fewer habitats. We calculated the average niche breadth index (*B*m) from all taxa of a sub-community to represent the habitat utilization, and the average *B*m values based on 100 bootstraps was used to plot each sub-community. We further classified the OTUs as generalists or specialists based on their occurrence and by using permutation algorithms as implemented in EcolUtils (Salazar, 2015). Specifically, the occurrences of OTUs in 1000 simulated OTU tables were individually calculated based on the *quasiswap* permutation algorithms. An OTU was defined as a generalist when the observed occurrence exceeded the upper 95% confidence interval (Supporting Information Appendix S2), and it was defined as a specialist when the observed occurrence fell below the lower 95% confidence interval (1000 permutations). The OTUs with an observed niche breadth within the 95% confidence interval were considered neutral taxa. We used this novel classification approach rather than an arbitrary cutoff for the *B*-value to avoid bias related to rare taxa, which may have been considered specialists simply because of their rarity.

*Community taxonomic composition*

We calculated the taxonomic community composition as the sum of the relative abundances of all OTUs belonging to a given taxonomic group (roughly to the phylum level based on assignments against the $PR^2$ database). Again, the taxonomic composition was estimated based on the average percentage of 100 bootstraps.

*Phylogenetic diversity*

We used Faith's PD index (Faith, 1992) to determine the phylogenetic alpha diversity. The PD index was estimated for each sub-community as the sum of the branch length associated with the phylogenetic tree using the package Picante (Kembel *et al.*, 2010). Pearson's correlation coefficient was calculated to test the relationships between simple environmental variables (latitude, temperature and chlorophyll *a* concentration) and the PD indices.

## Acknowledgements

We sincerely thank J. Hu for the CTD data, L. Wang for the pigment data and chief scientist M. Dai for the nutrient data and for organizing the cruises. We also thank the captain and crew of the R/V *Dongfanghong 2* for their help with sample collection. We thank V. Balagué and R. Massana at ICM-CSIC for their help with the 454 analyses and their general advice. This study was supported by the National Key Research and Development Program (No. 2016YFA0601201), the Natural Science Foundation of China (No. 41330961), the Doctoral Fund of the Ministry of Education of China (No. 20130121110031) and the Special Fund of the State Oceanic Administration of the People's Republic of China (No. GASI-03-01-02-03). The authors have no conflict of interest to declare.

## References

Alahuhta, J., Johnson, L.B., Olker, J., and Heino, J. (2014) Species sorting determines variation in the community composition of common and rare macrophytes at various spatial extents. *Ecol Complex* **20:** 61–68.

Alonso-Sáez, L., Díaz-Pérez, L., and Morán, X.A.G. (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environ Microbiol* **17:** 3766–3780.

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

Beisner, B.E., Peres-Neto, P.R., Lindström, E.S., Barnett, A., and Longhi, M.L. (2006) The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology* **87:** 2985–2891.

Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci U S A* **108:** 12776–12781.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7:** 335–336.

Caron, D.A., Peele, E.R., Lim, E.L., and Dennett, M.R. (1999) Picoplankton and nanoplankton and their trophic coupling in surface waters of the Sargasso Sea south of Bermuda. *Limnol Oceanogr* **44:** 259–272.

Cottenie, K. (2005) Integrating environmental and spatial processes in ecological community dynamics. *Ecol Lett* **8:** 1175–1182.

Countway, P.D., Gast, R.J., Savai, P., and Caron, D.A. (2005) Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. *J Eukaryot Microbiol* **52:** 95–106.

Debroas, D., Hugoni, M., and Domaizon, I. (2015) Evidence for an active rare biosphere within freshwater protists community. *Mol Ecol* **24:** 1236–1247.

Decelle, J., Martin, P., Paborstava, K., Pond, D.W., Tarling, G., Mahé, F., *et al.* (2013) Diversity, ecology and biogeochemistry of cyst-forming Acantharia (Radiolaria) in the oceans. *PLoS ONE* **8:** e53598.

Dray, S., Legendre, P., and Peres-Neto, P.R. (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Modell* **196:** 483–493.

Du, C., Liu, Z., Dai, M., Kao, S.J., Cao, Z., Zhang, Y., *et al.* (2013) Impact of the Kuroshio intrusion on the nutrient inventory in the upper northern South China Sea: insights from an isopycnal mixing model. *Biogeosciences* **10:** 6419–6432.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460–2461.

Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61:** 1–10.

Gao, Y., Arimoto, R., Duce, R.A., Lee, D.S., and Zhou, M.Y. (1992) Input of atmospheric trace elements and mineral matter to the Yellow Sea during the spring of a low-dust year. *J Geophys Res Atmos* **97:** 3767–3777.

Gong, J., Shi, F., Ma, B., Dong, J., Pachiadaki, M., Zhang, X., and Edgcomb, V.P. (2015) Depth shapes $\alpha$- and $\beta$-diversities of microbial eukaryotes in surficial sediments of coastal ecosystems. *Environ Microbiol* **17:** 3722–3737.

Guillou, L., Viprey, M., Chambouvet, A., Welsh, R.M., Kirkham, A.R., Massana, R., *et al.* (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to *Syndiniales* (*Alveolata*). *Environ Microbiol* **10:** 3349–3365.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., *et al.* (2013) The Protist Ribosomal Reference database ($PR^2$): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* **41:** D597–D604.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res* **21:** 494–504.

Hamilton, A.K., Lovejoy, C., Galand, P.E., and Ingram, R.G. (2008) Water masses and biogeography of picoeukaryote assemblages in a cold hydrographically complex system. *Limnol Oceanogr* **53:** 922–935.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497–506.

Hu, J., Kawamura, H., Hong, H., and Qi, Y. (2000) A review on the currents in the South China Sea: seasonal circulation, South China Sea warm current and Kuroshio intrusion. *J Oceanogr* **56:** 607–624.

Hubert, N., Calcagno, V., Etienne, R.S., and Mouquet, N. (2015) Metacommunity speciation models and their implications for diversification theory. *Ecol Lett* **18:** 864–881.

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Jouan Dufournel, I., Bronner, G., *et al*. (2013) Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc Natl Acad Sci U S A* **110:** 6004–6009.

Ichikawa, H., and Beardsley, R.C. (2002) The current system in the Yellow and East China Seas. *J Oceanogr* **58:** 77–92.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., *et al*. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26:** 1463–1464.

Kilias, E., Kattner, G., Wolf, C., Frickenhaus, S., and Metfies, K. (2014) A molecular survey of protist diversity through the central Arctic Ocean. *Polar Biol* **37:** 1271–1287.

Kim, T.W., Lee, K., Najjar, R.G., Jeong, H.D., and Jeong, H.J. (2011) Increasing N abundance in the northwestern Pacific Ocean due to atmospheric nitrogen deposition. *Science* **334:** 505–509.

Legendre, P., and Gauthier, O. (2014) Statistical methods for temporal and space-time analysis of community composition data. *Proc R Soc B* **281:** 20132728.

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., *et al*. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecol Lett* **7:** 601–613.

Levins, R. (1968) *Evolution in Changing Environments*. Princeton, New Jersey: Princeton University Press.

Liu, L., Yang, J., Yu, Z., and Wilkinson, D.M. (2015) The biogeography of abundant and rare bacterioplankton in lakes and reservoirs of China. ISME *J* **9:** 2068–2077.

Logares, R., Audic, S., Bass, D., Bittner, L., Boutte, C., Christen, R., *et al*. (2014) Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* **24:** 813–821.

Logares, R., Mangot, J.F., and Massana, R. (2015) Rarity in aquatic microbes: placing protists on the map. *Res Microbiol* **166:** 831–841.

Lozupone, C., and Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71:** 8228–8235.

Lynch, M.D.J., and Neufeld, J.D. (2015) Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* **13:** 217–229.

Martin, P., Allen, J.T., Cooper, M.J., Johns, D.G., Lampitt, R.S., Sanders, R., and Teagle, D.A.H. (2010) Sedimentation of acantharian cysts in the Iceland Basin: strontium as a ballast for deep ocean particle flux, and implications for acantharian reproductive strategies. *Limnol Oceanogr* **55:** 604–614.

Marzluff, J.M., and Dial, K.P. (1991) Life history correlates of taxonomic diversity. *Ecology* **72:** 428–439.

Massana, R. (2011) Eukaryotic picoplankton in surface oceans. *Annu Rev Microbiol* **65:** 91–110.

Massana, R., del Campo, J., Sieracki, M.E., Audic, S., and Logares, R. (2014) Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. ISME *J* **8:** 854–866.

McArdle, B.H., and Anderson, M.J. (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82:** 290–297.

Mojica, K.D.A., Huisman, J., Wilhelm, S.W., and Brussaard, C.P.D. (2016) Latitudinal variation in virus-induced mortality of phytoplankton across the North Atlantic Ocean. ISME *J* **10:** 500–513.

Monier, A., Worden, A.Z., and Richards, T.A. (2016) Phylogenetic diversity and biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. *Environ Microbiol Rep* **8:** 461–469.

Mouillot, D., Bellwood, D.R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., *et al*. (2013) Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol* **11:** e1001569.

Musat, N., Halm, H., Winterholler, B., Hoppe, P., Peduzzi, S., Hillion, F., *et al*. (2008) A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc. Natl Acad Sci U S A* **105:** 17861–17866.

Not, F., del Campo, J., Balagué, V., de Vargas, C., and Massana, R. (2009) New insights into the diversity of marine picoeukaryotes. *PLoS ONE* **4:** e7143.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., *et al*. (2014) vegan: community ecology package. R package version 2.2-0. http://CRAN.R-project.org/package=vegan/.

Östman, Ö., Drakare, S., Kritzberg, E.S., Langenheder, S., Logue, J.B., and Lindström, E.S. (2010) Regional invariance among microbial communities. *Ecol Lett* **13:** 118–127.

Painter, S., Patey, M., Tarran, G., and Torres-Valdés, S. (2014) Picoeukaryote distribution in relation to nitrate uptake in the oceanic nitracline. *Aquat Microb Ecol* **72:** 195–213.

Pandit, S.N., Kolasa, J., and Cottenie, K. (2009) Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology* **90:** 2253–2262.

Pedrós-Alió, C. (2006) Marine microbial diversity: can it be determined?. *Trends Microbiol* **14:** 257–263.

Pedrós-Alió, C. (2012) The rare bacterial biosphere. *Ann Rev Mar Sci* **4:** 449–466.

Pendleton, R.M., Hoeinghaus, D.J., Gomes, L.C., and Agostinho, A.A. (2014) Loss of rare fish species from tropical floodplain food webs affects community structure and ecosystem multifunctionality in a mesocosm experiment. *PLoS ONE* **9:** e84568.

Pester, M., Bittner, N., Deevong, P., Wagner, M., and Loy, A. (2010) A "rare biosphere" microorganism contributes to sulfate reduction in a peatland. ISME *J* **4:** 1591–1602.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) Fasttree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26:** 1641–1650.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al*. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41:** 590–596.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Reeder, J., and Knight, R. (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat Methods* **7:** 668–669.

Salazar, G. (2015) EcolUtils: utilities for community ecology analysis. R package version 0.1. https://github.com/GuillemSalazar/EcolUtils.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Siqueira, T., Bini, L.M., Roque, F.O., Marques Couceiro, S.R., Trivinho-Strixino, S., and Cottenie, K. (2012) Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. *Ecography* **35:** 183–192.

Steinhoff, T., Friedrich, T., Hartman, S.E., Oschlies, A., Wallace, D.W.R., and Körtzinger, A. (2010) Estimating mixed layer nitrate in the North Atlantic Ocean. *Biogeosciences* **7:** 795–807.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.W., and Richards, T.A. (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol* **19:** 21–31.

Tucker, C.M., Cadotte, M.W., Carvalho, S.B., Davies, T.J., Ferrier, S., Fritz, S.A., *et al.* (2016) A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biol Rev* doi:10.1111/brv.12252.

Unrein, F., Gasol, J.M., Not, F., Forn, I., and Massana, R. (2014) Mixotrophic haptophytes are key bacterial grazers in oligotrophic coastal waters. ISME *J* **8:** 164–176.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., *et al.* (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* **348:** 1261605.

Venrick, E.L., McGowan, J.A., and Mantyla, A.W. (1973) Deep maxima of photosynthetic chlorophyll in the Pacific Ocean. *Fish Bull* **71:** 41–52.

Vergin, K.L., Done, B., Carlson, C.A., and Giovannoni, S.J. (2013) Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. *Aquat Microb Ecol* **71:** 1–13.

Wu, W., Huang, B., and Zhong, C. (2014) Photosynthetic picoeukaryote assemblages in the South China Sea from the Pearl River estuary to the SEATS station. *Aquat Microb Ecol* **71:** 271–284.

Wu, W., Wang, L., Liao, Y., and Huang, B. (2015) Microbial eukaryotic diversity and distribution in a river plume and cyclonic eddy-influenced ecosystem in the South China Sea. *MicrobiologyOpen* **4:** 826–840.

Yeh, Y.C., Peres-Neto, P.R., Hunag, S.W., Lai, Y.C., Tu, C.Y., Shiah, F.K., *et al.* (2015) Determinism of bacterial metacommunity dynamics in the southern East China Sea varies depending on hydrography. *Ecography* **38:** 198–212.

Yuasa, T., and Takahashi, O. (2016) Light and electron microscopic observations of the reproductive swarmer cells of nassellarian and spumellarian polycystines (Radiolaria). *Eur J Protistol* **54:** 19–32.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. *J Comput Biol* **7:** 203–214.

## Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Table S1.** Sample list and environmental variables in each sampling location. Depth (m); temperature (°C); salinity (psu); Chl *a*, chlorophyll *a* ($\mu g\ l^{-1}$); $NO_x$, nitrate and nitrite ($\mu M$); $PO_4$, phosphate ($\mu M$); and MLD, mixed layer depth (m). Samples are labelled with the prefix CC01 for the summer cruise, CC02 for the winter cruise, CC03 for the fall cruise, and CC04 for the spring cruise.

**Table S2.** Top 20 OTUs based on the average relative contributions (%) of sequence abundances in all 50 samples.

**Table S3.** Sequence information (number of removed singletons and final quality-checked reads), diversity index (Shannon diversity index *H'* and Chao 1 index with a threshold of 4958 reads), and average read and OTU numbers for abundant and rare taxa based on 100 subsamples (with 4958 reads). OTU, operational taxonomic unit at 99% similarity level. Samples are labelled with the prefix CC01 for the summer cruise, CC02 for the winter cruise, CC03 for the fall cruise and CC04 for the spring cruise.

**Fig. S1.** Ordination biplot of the principal component analysis of environmental variables in the surface (A) and subsurface (B) layers. T, temperature; S, salinity; Chl, chlorophyll *a*; $NO_x$, nitrate and nitrite; $PO_4$, phosphate; MLD, mixed layer depth.

**Fig. S2.** Relative contributions of habitat generalists and specialists in the abundant and rare sub-communities in the surface and subsurface layers. The results are based on the average of 100 bootstraps.

**Fig. S3.** Taxonomic composition patterns for relative sequence abundances at the phylum level for abundant (A and C) and rare (B and D) sub-communities. Samples are labelled with the prefix CC01 for the summer cruise, CC02 for the winter cruise, CC03 for the fall cruise and CC04 for the spring cruise.

**Appendix S1.** QIIME script for picking OTUs, multiple rarefactions (for bootstrap analyses) and taxonomy assignments.

**Appendix S2.** R script for the 3-way PERMANOVA and definition of generalists and specialists.