


The diversity and biogeography of abundant and rare intertidal marine microeukaryotes explained by environment and dispersal limitation

Wenjing Zhang ^{1,2*}, Yongbo Pan,¹ Jun Yang,³ Huihuang Chen,³ Bridget Holohan,² Jamie Vaudrey,² Senjie Lin^{1,2} and George B. McManus^{2**}

¹State Key Laboratory of Marine Environmental Science, Marine Biodiversity and Global Change Research Center, Xiamen University, Xiamen 361005, People's Republic of China.

²Department of Marine Sciences, University of Connecticut, Groton, CT 06340, USA.

³Aquatic EcoHealth Group, Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, People's Republic of China.

Summary

Benthic microeukaryotes are key ecosystem drivers in marine sandy beaches, an important and dynamic environment; however, little is known about their diversity and biogeography on a large spatial scale. Here, we investigated the community composition and geographical distributions of benthic microeukaryotes using high-throughput sequencing of the 18S rRNA gene and quantified the contributions of environmental factors and spatial separation on the distribution patterns of both rare and abundant taxa. We collected 36 intertidal samples at 12 sandy beaches from four regions that spanned distances from 0.001 to 12,000 km. We found 12,890 operational taxonomic units (OTUs; 97% sequence identity level) including members of all eukaryotic super-groups and several phyla of uncertain position. Arthropoda and Diatomeae dominated the sequence reads in abundance, but Ciliophora and Discoba were the most diverse groups across all samples. About one-third of the OTUs could not be definitively classified at a similarity level of 80%, supporting the view that a

large number of rare and minute marine species may have escaped previous characterization. We found generally similar geographical patterns for abundant and rare microeukaryotic sub-communities, and both showed a significant distance–decay similarity trend. Variation partitioning showed that both rare and abundant sub-communities exhibited a slightly stronger response to environmental factors than spatial (distance) factors. However, the abundant sub-community was strongly correlated with variations in spatial, environmental and sediment grain size factors (66% of variance explained), but the rare assemblage was not (16%). This suggests that different or more complex mechanisms generate and maintain diversity in the rare biosphere in this habitat.

Introduction

Microbial eukaryotes are important components of marine ecosystems, with extremely diverse taxa in multiple trophic modes including autotrophy, mixotrophy, predation, parasitism, symbiosis, osmotrophy and saprotrophy (Zubkov and Tarran, 2008; Worden *et al.*, 2015). Thus, understanding the diversity and biogeographical patterns of microeukaryotes is critical to understanding ecological processes and mechanisms that underlie and maintain ecosystem function at both regional and global scales (Hanson *et al.*, 2012). However, our understanding of the composition of marine microbial communities and the ecological factors that determine community structure at a global scale is very limited (Nunes-Alves, 2015), especially in the rare microbial biosphere of benthic environments (Logares *et al.*, 2014; Lynch and Neufeld, 2015). Additionally, understanding microbial ecosystems is critical for predicting ecosystem function as global environments change (Margules and Pressey, 2000; Worm *et al.*, 2006). Until very recently, both microbes and marine systems were hardly featured at all in many textbooks on biogeography, which are still mostly concerned with terrestrial and freshwater macroscopic organisms. Thus, the topic of the biodiversity and distribution patterns of microeukaryotes in

Received 3 May, 2017; revised 15 August, 2017; accepted 26 August, 2017. *For correspondence. *E-mail zhangwenjing@xmu.edu.cn. Tel./Fax: (+86)592-2182374. **E-mail george.mcmanus@uconn.edu.

intertidal sandy beaches is quite a new area in biogeography.

Advances in high-throughput sequencing (HTS) combined with metabarcoding approaches have improved our ability to assess the biodiversity of microeukaryotes and address ecologically relevant questions on an unprecedented scale (Bik *et al.*, 2012). Recently, HTS techniques have been successfully used to examine the composition and dynamics of microbial communities in different ecosystems, including surface and deep marine waters (Edgcomb *et al.*, 2011a; Pernice *et al.*, 2015; Yu *et al.*, 2015), marine sediments (Gong *et al.*, 2015), lakes and reservoirs (Simon *et al.*, 2015; Liu *et al.*, 2017), soils (Bates *et al.*, 2013) and estuaries (Lallias *et al.*, 2015). Unexpectedly, high microeukaryotic diversity has been discovered in all of these studies. For example, de Vargas *et al.* (2015) estimated the diversity of small eukaryotes (less than about 2 mm in size) in the sunlit ocean to be ~150,000 OTUs, much higher than previously thought (de Vargas *et al.*, 2015).

Most microbial communities consist of a large number of species partitioned into a few abundant forms and many that are represented by only a few individuals (Sogin *et al.*, 2006; Logares *et al.*, 2014; Liu *et al.*, 2015). The latter, referred to as the 'rare biosphere', are likely subject to different controlling factors than the more abundant species, and an important effort of a growing body of studies in recent years has been directed toward understanding the dynamics of the rare biosphere because rare taxa can have important ecological roles, serving as nearly limitless reservoirs of genetic and functional diversity (Pedrós-Alió, 2012; Lynch and Neufeld, 2015). Because microbes are not dispersal-limited in the same way that large organisms are, the geographical distribution of rare microbial taxa is important for understanding overall microeukaryotic diversity (Telford *et al.*, 2006; Logares *et al.*, 2014). Although the existence of microbial biogeographic patterns is now well established in microbial ecology, there is limited understanding of the mechanisms governing the relative influences of environmental deterministic and stochastic processes in driving these patterns (Hanson *et al.*, 2012; Liu *et al.*, 2015). Recent studies showed that the spatial and temporal patterns of rare and abundant taxa in marine bacteria and archaea are usually different in coastal ecosystems (Gobet *et al.*, 2012; Hugoni *et al.*, 2013). However, some studies found similar patterns for the abundant and rare assemblages in aquatic ecosystems (Gong *et al.*, 2015; Liu *et al.*, 2015). At this point, it is not known whether the microeukaryotic community composition of the rare biosphere is influenced more by spatial factors (ability to disperse over great distances to new habitats) or environmental factors (ability to grow under a range of environmental conditions). Intertidal sandy beaches provide an ideal system to address these questions at a large

geographical scale in a ubiquitous, complex and dynamic habitat.

Sandy beaches comprise 75% of the world's unfrozen shorelines, providing key ecosystem services, such as water purification, storm buffering, nutrient cycling and nursery habitats for resource species, and have also become important for economic development and cultural use (Nel *et al.*, 2014). This habitat is very dynamic in space and time, due to the interaction of sand, waves and tides (Short, 1996). A large number of microbes and small animals (meiofauna) have adapted to the changing environmental conditions of sandy beaches (McLachlan and Brown, 2006). In addition to these residents, some species descend onto the beach from inland on the falling tide, while a variety of species move up over the beach from the sub-littoral zone on the rising tide. The open ecosystem of the sandy beach is created by the interaction of these components in a trophic network, which exchanges materials with both sea and land (McLachlan and Brown, 2006). The environmental quality of sandy beaches is important to the health and safety of beach users and the economic prosperity of coastal communities (Zhang *et al.*, 2015). Over the past decades, however, the microbial species diversity of sandy beaches has only been explored at very low resolution levels (i.e., with ordinary microscopy). Thus, little is known about the diversity patterns and relative importance of spatial and environmental factors for the microeukaryotic community. In fact, the information and hypotheses about global patterns of species richness on sandy beaches have mainly focused on macro- and meio-fauna in recent decades (McArdle and McLachlan, 1992; Defeo and McLachlan, 2005; Harris *et al.*, 2014; Nel *et al.*, 2014), and very few studies have quantified the rare microbial species (Fonseca *et al.*, 2014). New methods using HTS and bioinformatic tools allow us to quantify microeukaryotic diversity of this important and unique habitat in a way that encompasses both the abundant and rare members of the ecosystem at a meaningful sequencing depth.

Here, we present the first comprehensive study of large-scale distribution of microeukaryotic benthic communities in sandy beaches along a latitudinal gradient ranging from 24 to 42°N at a continental scale (North America and Asia). We determine and compare the patterns in alpha and beta diversities of both abundant and rare microeukaryotic benthic taxa, characterized by Illumina high-throughput sequencing (HTS) of the 18S rRNA gene, across 12 intertidal stations with three replicates at each station. In addition, we assess whether spatial and environmental factors (including sediment grain size) explained the geographical patterns of both abundant and rare microeukaryotic benthic sub-communities.

Results

General patterns of species richness

Our study system consists of 12 sandy beaches across three different regions [Xiamen (XM), Zhoushan (ZS) and Qingdao (QD) cities] of eastern China and one in Connecticut (CT) of northeastern United States (Fig. 1 and Table S1 in the Supporting Information). The four regions sampled span large gradients in terms of climate and latitude, spanning a geographical distance up to 12,000 km. The benthic microeukaryotes (0.22–200 μm) in 36 sediment samples, based on HTS of 18S rDNA amplicons, yielded 1,232,604 clean reads and 12,890 OTUs at 97% similarity level after singleton OTUs were removed (Table S2 in the Supporting Information). In the whole dataset, 171 (1.3%) OTUs with 617,908 sequences (50.1%) were classified as abundant, whereas 7553 (58.6%) OTUs with 37,259 (3.0%) sequences were rare (Table S2 in the Supporting Information).

The number of OTUs varied from 1679 (sample GR.B3) to 3994 (ZS.B2) per sample (mean = 2748, s.e. = 104, $n = 36$) (Tables S3 in the Supporting Information). The rarefaction curves (OTUs detected vs. sequencing effort) were not saturated for most individual samples (Fig. 2A), but the total number of OTUs in all samples (12,890) was roughly equivalent to the number estimated by abundance-based

richness estimators such as Chao 1 ($12,895 \pm 2$) and ACE (abundance-based coverage estimator, $12,927 \pm 56$) (Table S3 in the Supporting Information). Both the estimated species accumulation curves (Fig. 2C) and extrapolated species richness indices (Chao 1 and ACE) indicated that the majority of the microeukaryotic taxa had been well recovered from the 36 studied sites. The Shannon index, providing an estimate of alpha diversity in each sample, ranged from 3.27 to 6.73 with a mean of 5.50 ± 0.13 (s.e.) (Table S3 in the Supporting Information) and did not differ across four regions ($P > 0.05$), with 5.51 ± 0.31 in CT, 5.61 ± 0.13 in QD, 5.31 ± 0.34 in ZS and 5.52 ± 0.19 in XM. Other alpha diversity indices (including OTU number, ACE, Chao 1, Simpson and Pielou) also did not show a significant difference within and across regions (ANOVA, $P > 0.05$, Fig. S1 in the Supporting Information).

A total of 4388 OTUs (34.0%) could not be assigned to specific taxa based on $>80\%$ sequence similarity (on average, 9.2% of sequences per sample) and were lumped here as unclassified (Fig. 3B). The remaining 8502 assignable OTUs were classified into 43 deep-branching lineages covering the full spectrum of cataloged eukaryotic diversity among seven recognized super-groups (Alveolata, Amoebozoa, Archaeplastida, Excavata, Opisthokonta, Rhizaria and Stramenopiles) and multiple lineages of uncertain placement (Apusomonadida, Centrohelida, Cryptophyta

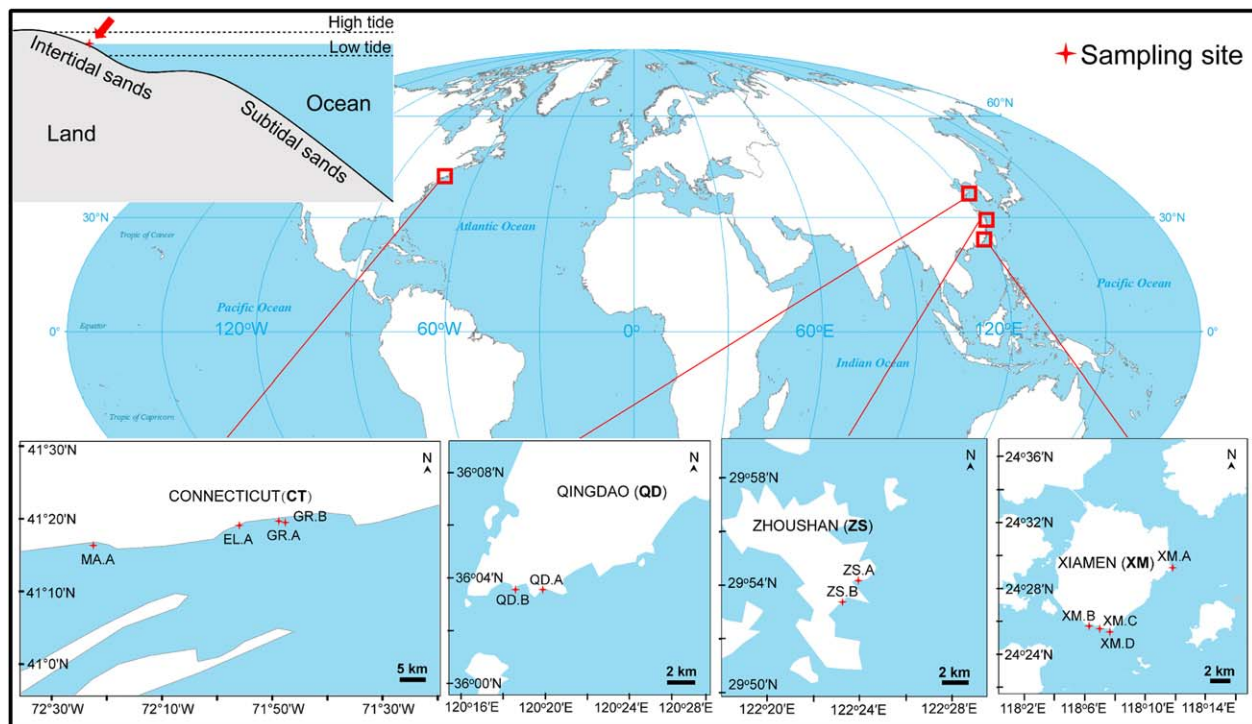


Fig. 1. Location of the 12 intertidal sampling stations in four regions (CT from USA and QD, ZS and XM from China).

For each of the sampling stations, three replicate samples were taken approximately 1 m apart from each other on the falling tide, which just inundated the sandy sediment. Map was performed in ARCGIS 10.1. [Colour figure can be viewed at wileyonlinelibrary.com]

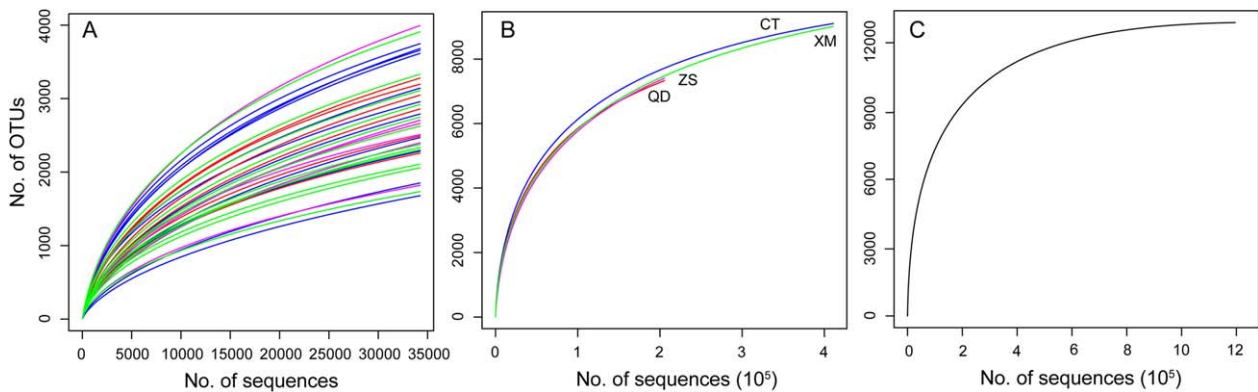


Fig. 2. Rarefaction curves of similarity-based operational taxonomic unit (OTUs) at 97% sequence identity threshold.

A. The individual intertidal benthic microeukaryotes samples.

B. The combined sets of four regions samples (12 replicate samples in both CT and XM and 6 replicate samples in both QD and ZS).

C. The combined set of 36 samples. Four sampling regions are shown in different colours: CT (blue), QD (red), ZS (pink) and XM (green).

[Colour figure can be viewed at wileyonlinelibrary.com]

and Haptophyta) (Fig. 3B). Each of the major lineages was represented by more than 1000 sequences, with nine lineages (Chloroplastida, Arthropoda, Gastrotricha, Nematoda, Platyhelminthes, Ciliophora, Dinoflagellata, Cercozoa and Diatomea) dominating, each comprising more than 4% (49,304) of the sequences and together representing 65% of all sequences. Among these, Arthropoda and Diatomea were the most dominant, comprising 13.8% and 14.1% of the sequence reads, respectively (Fig. 3C). Eight 'hyperdiverse' lineages (Chloroplastida, Discoba, Arthropoda, Fungi, Ciliophora, Dinoflagellata, Cercozoa and Diatomea) each contained more than 200 OTUs, together representing 35% of the total OTUs (Fig. 3B). Among these, Ciliophora and Discoba had the most diversity, contributing 6.4% and 5.7% of the OTU richness, respectively. Only 30 out of the 12,890 total OTUs were found in all 36 samples, and nearly half of the OTUs (6290, 48.8%) were found in fewer than five samples (Fig. S2 in the Supporting Information).

Geographical patterns

At higher taxonomic level, the microeukaryote communities showed a distinct biogeographical distribution pattern (Fig. 3A). This was true at the OTU level for the whole dataset as well as both the abundant and rare sub-communities (Fig. 4). The relative abundance of OTUs was significantly correlated with their local occupancy (Fig. S3 in the Supporting Information) and only 27.9% microeukaryotic OTUs (3598) were shared among the four regions (Fig. S4 in the Supporting Information). Additionally, the four sampling regions were significantly separated at $P = 0.001$ for the microeukaryotic communities, and the global ANOSIM R statistics among the four groups was 0.671, 0.626 and 0.614 for the entire community, the abundant sub-community and the rare sub-community, respectively, indicating significant differences over this

large spatial range (Table 1). It is notable that the Connecticut and Qindao communities were more similar to each other than other regions (Table 1, R values lower than others; Fig. 4), likely due to their similar temperate latitude. Spearman correlograms comparing Bray–Curtis community similarity vs. geographic distances among samples indicated significant negative correlations for the entire community and both the abundant and rare sub-communities, with a correlation coefficient of -0.437 ($P < 0.01$), -0.444 ($P < 0.01$) and -0.346 ($P < 0.01$), respectively (Figs 5 and S5 in the Supporting Information). This means that the similarity in community composition between any two regions or sites decreased with increasing geographical distances. However, we did not find a significant latitudinal diversity gradient over the nearly 17 degree range of our samples (Fig. S1 in the Supporting Information; roughly from subtropical to temperate regions).

Relationships between spatial, environmental factors and microeukaryotic community

The longest gradient lengths of detrended correspondence analysis (DCA) on the abundant and rare sub-community data sets were 2.50 and 4.32, respectively, indicating that redundancy analysis (RDA) was suitable for the abundant sub-community, while canonical correspondence analysis (CCA) was more suitable for the rare sub-community. The RDA ordination showed that six environmental variables [temperature, salinity, pH, total nitrogen (TN), total phosphorus (TP) and phosphate phosphorus ($\text{PO}_4\text{-P}$)], six spatial variables [PCNM1–PCNM6 (principal coordinates of neighbour matrices)] and four grain size variables (S1, S2, S3 and S5) were significantly related to changes in the abundant sub-community ($P < 0.05$; Fig. S6 in the Supporting Information). The CCA ordination revealed that the rare

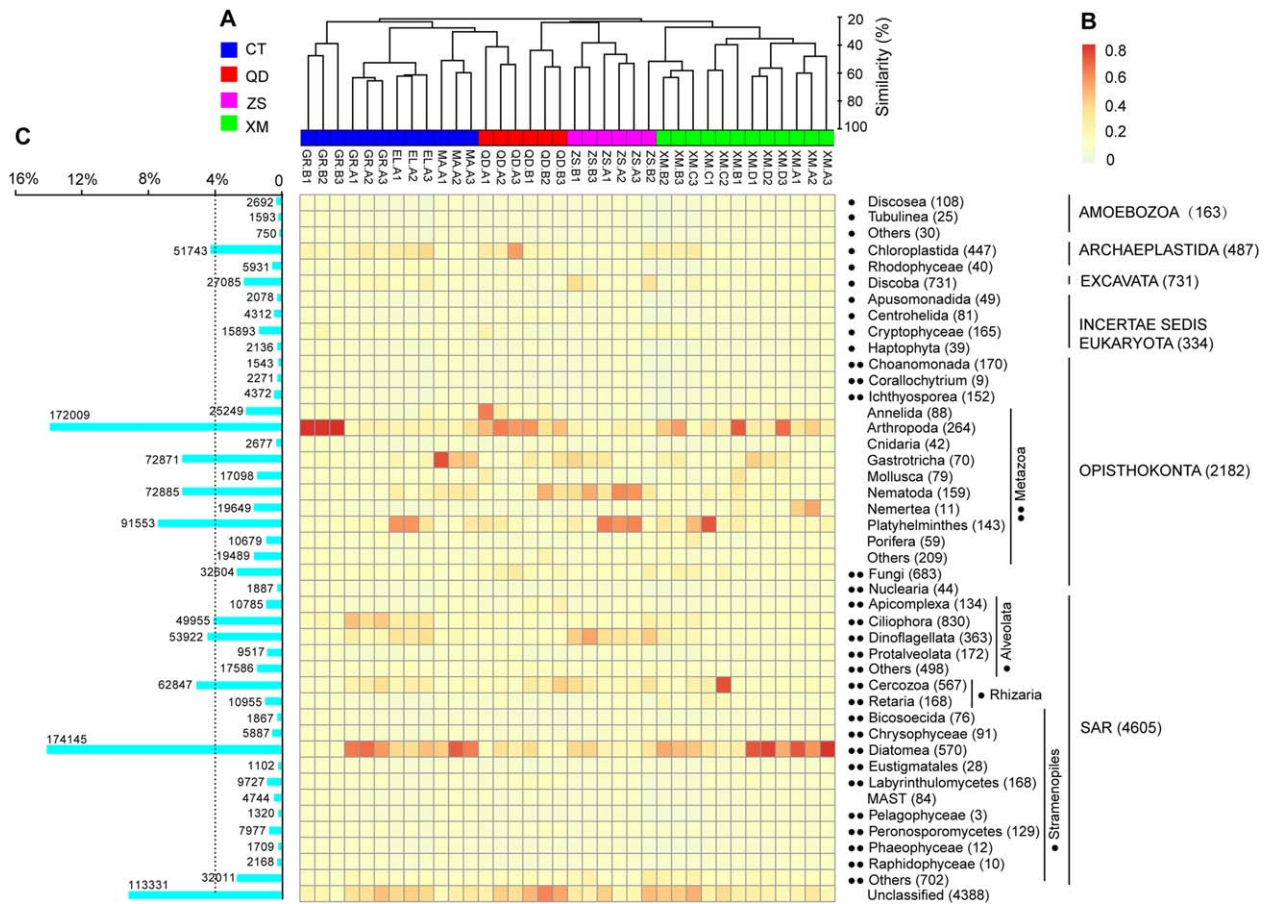


Fig. 3. The distribution and relative abundance of the intertidal benthic microeukaryotes. The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold. **A.** Dendrogram cluster for 36 samples from four regions based on Bray–Curtis similarity. The colour coding at the branch end based on their four sampling regions. **B.** Heatmap showing the relative abundance of the 43 lineages within each sample. The OTU numbers (species richness) are given in parentheses for each lineage, while the sequence data were normalized by range-scaling each class (0–1) for each sample. The lineages were classified mainly based on Adl and colleagues (2012); each of the lineages that contained more than 1000 sequence reads within entire dataset was individually presented, while the rare lineages (fewer than 1000 sequence reads) and the undetermined lineages (sequence similarity >80%) merged as ‘others’. Sequence reads are ‘unclassified’ when sequence similarity to a reference sequence is <80%. A single dot in front of the taxon indicates the first rank classification of the protists and multicellular groups, while double dots indicate the second rank classification. **C.** Bar graph showing the relative abundance of sequence reads in the 43 branching lineages within the entire dataset (metacommunity including 36 local communities). [Colour figure can be viewed at wileyonlinelibrary.com]

sub-communities have a slightly different pattern in that only five environmental variables (temperature, salinity, pH, TN and ORG), five spatial variables (PCNM1–PCNM5) and four grain size variables (S1, S2, S3 and S5) were significantly related to changes in the sub-community composition ($P < 0.05$; Fig. S6 in the Supporting Information).

Variation partitioning showed that distance, environmental factors and grain size together explained 54% of the entire community variation, including 66% for the abundant sub-community, but only 16% of variation for the rare sub-community (Fig. 6). Among all of these communities, the relative contribution of environmental plus grain size factors was slightly larger than that of spatial (distance)

factors. We further evaluated the variation of environmental and grain size factors and found that the variation of environmental factors was larger than that of grain size factors (Fig. 6). When the three kinds of variables were independently evaluated, the pure effects of spatial factors were larger than those of pure environmental or pure grain size factors (Fig. S7 in the Supporting Information). These results indicate that environmental factors were critical to benthic microeukaryotic communities, especially the abundant sub-communities. Because environmental factors were most likely also related to spatial (distance) factors (e.g., due to latitudinal, geological or land use differences in the different regions), they were also major drivers of the

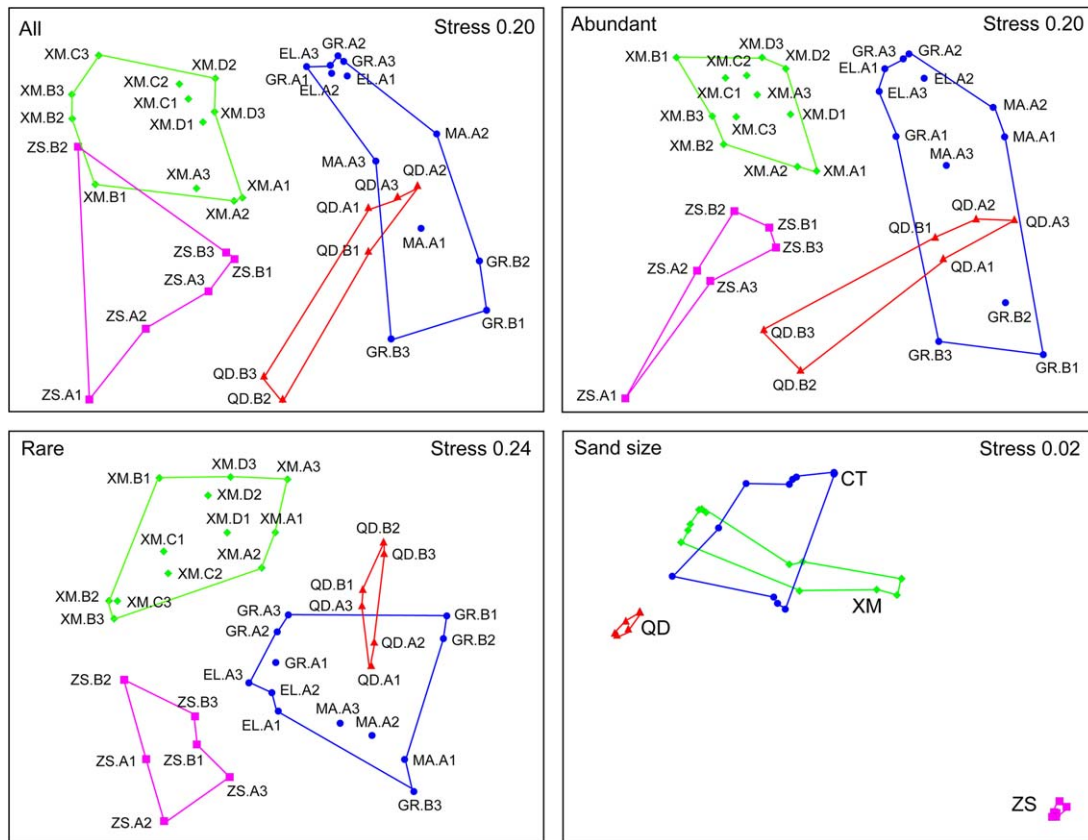


Fig. 4. Non-metric multidimensional scaling ordinations (nMDS) for intertidal benthic microeukaryote communities and sand size composition from 36 samples.

All: all microeukaryotic taxa; abundant: abundant taxa; rare: rare taxa. Four sampling regions are shown in different colours: CT (blue), QD (red), ZS (pink) and XM (green). The lines around the sites of each sampling regions were manually drawn and have no special meaning. [Colour figure can be viewed at wileyonlinelibrary.com]

spatial differentiation observed, although the contribution of each environmental driver differed somewhat between regions.

Despite lower correlation coefficients for rare sub-communities and environmental and distance factors, compared to the abundant sub-communities, Mantel and partial Mantel tests further confirmed that both abundant and rare sub-communities were primarily governed by environmental plus sand grain size variables, rather than distance factors (Table 2).

Discussion

Diversity and distribution

In our samples from temperate and subtropical intertidal marine sandy sediments, we found members of all eukaryotic super-groups as well as some phyla of uncertain position (Fig. 3). Alveolata, Opisthokonta and Stramenopiles were the most diverse and abundant groups, as observed in surface sediments from other coastal ecosystems (Adl *et al.*, 2012; Gong *et al.*, 2015).

Although a 200- μm pore size sieve was used to remove large metazoans, we found that the most abundant single group was the Opisthokonta, which comprised 44.4% of all sequence reads. This group was dominated by metazoans, which accounted for 40.9% of all sequences. Some large-sized organisms (esp. elongated species) can get through the 200- μm mesh. Also, the presence of smaller life stages (eggs, spores or larvae) of large-sized organisms may have contributed to the assemblage (Liu *et al.*, 2017). The top four metazoan phyla, in terms of both OTUs and sequences, were Arthropoda, Platyhelminthes, Nematoda and Gastrotricha as seen in previous metabarcoding analyses of marine meiofaunal biodiversity (Fonseca *et al.*, 2010; 2014; Lallias *et al.*, 2015). Additionally, a great diversity of unicellular Opisthokonts (e.g., choanomonads, ichthyosporeans and fungi; 8.2% of all OTUs) were recovered although their abundances were very low (3.5% of total sequences). For example, the marine fungi presented high diversity (683 OTUs, 5.3% of the total OTUs) but low abundance (2.6% of the total sequences) in all of the sampling sites, similar to other

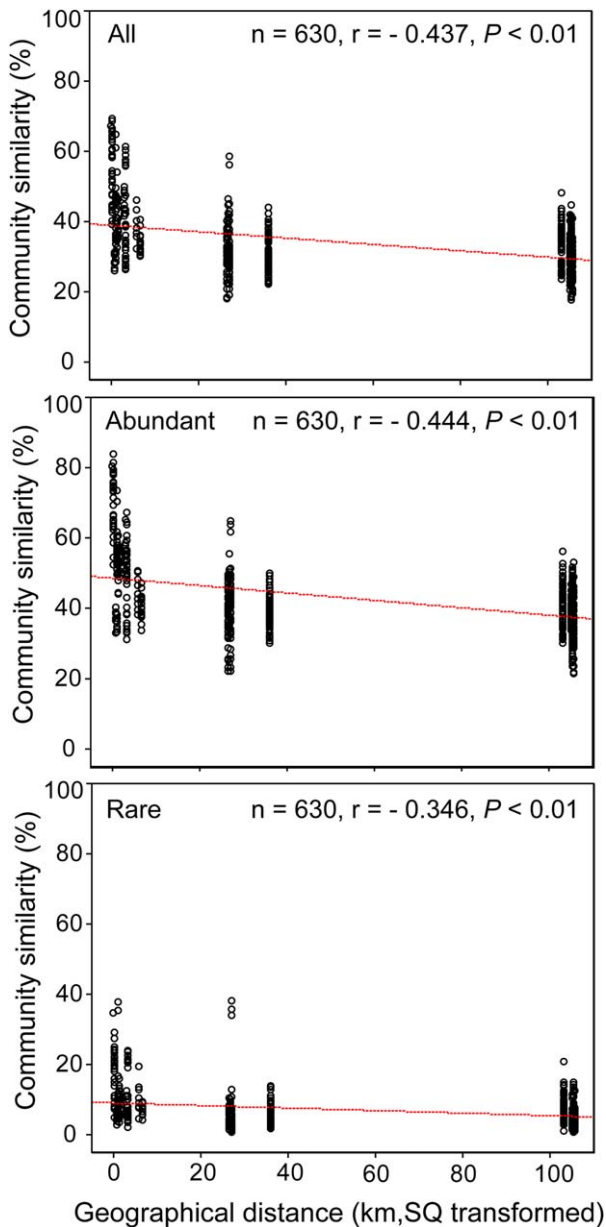


Fig. 5. Spearman's rank correlations between the Bray–Curtis similarity of microeukaryotic community and geographical distance between sampling sites (n is the number of comparison and P values are indicated).

All: all microeukaryotic taxa; abundant: abundant taxa; rare: rare taxa. Geographical distances were square-root (SQ) transformed. [Colour figure can be viewed at wileyonlinelibrary.com]

studies that described low proportions of fungal sequences in surface marine waters (0.8%) (Massana and Pedrós-Alió, 2008), deep anoxic waters (1–4%) (Edgcomb *et al.*, 2011a) and coastal surface sediments (3.5%) (Gong *et al.*, 2015). This contrasts with deep subsurface sediments, where fungi appear to dominate the microeukaryotic community (Edgcomb *et al.*, 2011b). The high diversity of marine fungi suggests that they may be associated with

Table 1. Analysis of similarities (ANOSIM) of all, abundant and rare intertidal-benthic microeukaryotic communities among four different regions.

Pairwise tests	All		Abundant		Rare	
	R	P	R	P	R	P
CT vs. QD	0.304*	0.013	0.263*	0.023	0.170	0.053
CT vs. ZS	0.616**	<0.001	0.546**	<0.001	0.618**	<0.001
CT vs. XM	0.702**	<0.001	0.621**	<0.001	0.648**	<0.001
QD vs. ZS	0.807**	0.002	0.713**	0.002	0.920**	0.002
QD vs. XM	0.828**	<0.001	0.869**	<0.001	0.715**	<0.001
ZS vs. XM	0.762**	<0.001	0.811**	<0.001	0.734**	<0.001
Global test	0.671**	<0.001	0.626**	<0.001	0.614**	<0.001

The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold.

The ANOSIM results are generated using the Bray–Curtis similarity data.

* $P < 0.05$.

** $P < 0.01$.

the recycling of nutrients in sandy beaches, as recent reports showed that the marine fungi are correlated with total organic carbon, nitrate, sulfide and dissolved inorganic carbon, in anoxic marine pelagic environments and especially at the seafloor (Richards *et al.*, 2012; Orsi *et al.*, 2013; Worden *et al.*, 2015). Non-metazoan and non-fungal opisthokonts were also diverse although a large number of them are considered to be rare taxa, which are poorly known in many natural environments (del Campo *et al.*, 2015). An alternative explanation for the high diversity/low abundance pattern is that opisthokont microeukaryote cells or dispersal stages (e.g., fungal spores) are ubiquitous due to lack of dispersal limitation, but not always active.

The Stramenopiles were the second most dominant group, represented mainly by Diatomea. Diatoms are eukaryotic unicellular microalgae present in all types of aquatic ecosystems (Cermeno and Falkowski, 2009). Over one-third of the total 30 cosmopolitan OTUs in our data set were diatoms, suggesting high dispersal capability and/or ecological success of these species in diverse environments. Diatomea were also the dominant group observed in European coastal sediments (Massana *et al.*, 2015). MASTs are a group of heterotrophic marine Stramenopiles that are known mostly from environmental DNA surveys of both oxic and anoxic systems. They appear to be bacterivores and comprise a widespread basal branch of the Stramenopiles (Logares *et al.*, 2014) that may have significant role in marine plankton systems (Massana *et al.*, 2014). Our study recovered 84 MAST OTUs with 4744 sequences. They were minor components in all of our samples (0.38% of total sequences).

Alveolata sequences were abundant in our study, comparable to results of clone library analyses of microeukaryotes in other marine habitats, for example, anoxic deep waters (Edgcomb *et al.*, 2011a), the water column (Pernice *et al.*,

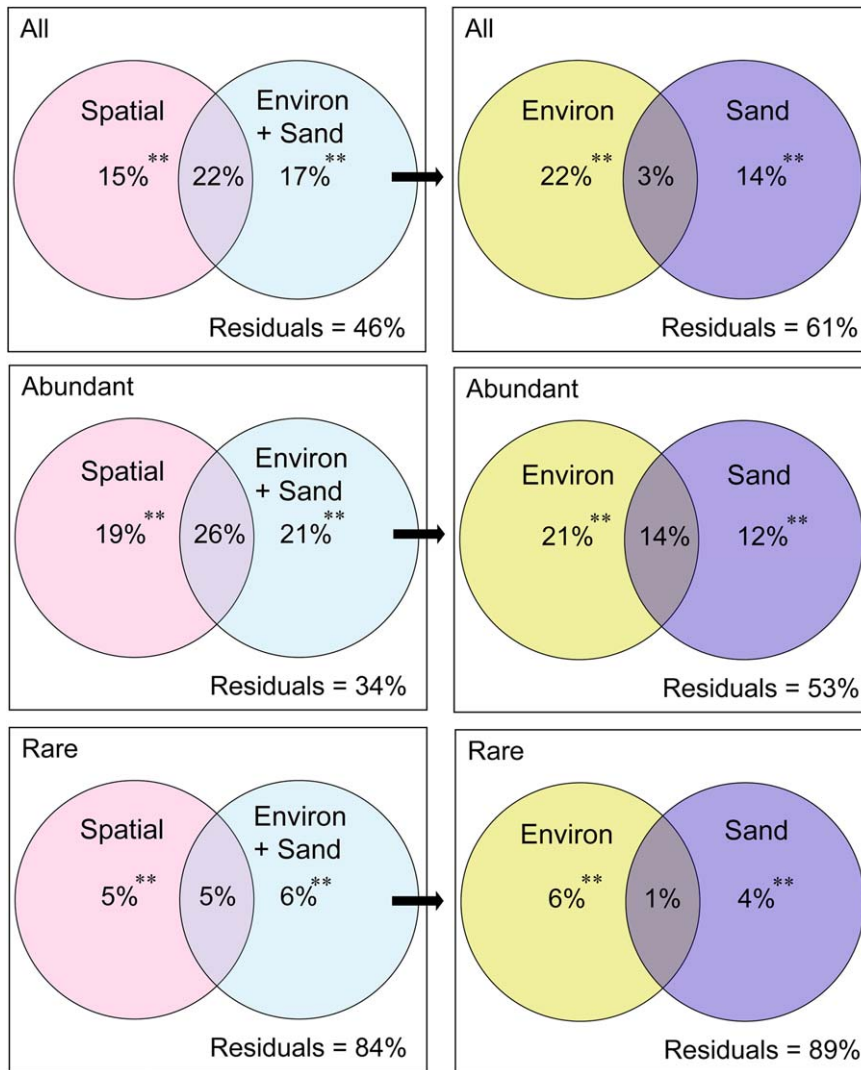


Fig. 6. Venn diagram of variation partitioning, showing the effects of spatial, environmental and sand grain size variables on the community composition of benthic microeukaryotes. Values indicate the percentage of variation explained by each fraction, including pure, shared explained and unexplained variability. Forward selection procedures were used to select the best subset of spatial, environmental and sand grain size variables explaining community variation, respectively. ANOVA permutation tests were calculated on the variation explained by each set without the effect of the other. Significance levels are as follows: * $P < 0.05$ and ** $P < 0.01$. All: all microeukaryotic taxa; abundant: abundant taxa; rare: rare taxa. [Colour figure can be viewed at wileyonlinelibrary.com]

2013) and abyssal seafloor sediments (Scheckenbach *et al.*, 2010).

About 34% of eukaryotic ribosomal diversity in our data set could not be classified because it did not closely match any reference barcode in the SILVA 115 database and thus could not be assigned to any of the known eukaryotic

Table 2. Mantel and partial Mantel tests for the correlation between community similarity and spatial, environmental and sand size variables using Pearson's coefficient.

Effects of	Control for	Abundant subcommunity	Rare subcommunity
Environ + sand		0.677**	0.532**
Spatial		0.638**	0.484**
Environ + sand	Spatial	0.483**	0.359**
Spatial	Environ + sand	0.402**	0.263**

The significances are tested based on 999 permutations.
** $P < 0.01$.

super-groups. However, this unclassified diversity represented only a small proportion (about 9.2%) of total sequences because a large proportion of these OTUs were rare. Based on our knowledge, there are two possible explanations for the large number of unclassified OTUs. First, some of these may correspond to divergent rDNA pseudogenes, known to exist in eukaryotes (Márquez *et al.*, 2003; Pan *et al.*, 2016), or sequencing artifacts (Decelle *et al.*, 2014). Second, it is more likely to be due to limited diversity coverage of the current V9 database of 18S rDNA sequences and alignments used to make taxon assignments (de Vargas *et al.*, 2015; Harder *et al.*, 2016). It has been estimated that between one-third and two-thirds of all marine species, corresponding primarily to rare and minute taxa, have escaped previous characterization and are thus not represented in public databases (Appeltans *et al.*, 2012; de Vargas *et al.*, 2015).

Rarefaction curves showed that our sequencing effort appeared satisfactory at a global level, indicating that most

of the abundant taxa present in the 12 sandy beaches have been retrieved by our sampling. This gives an estimate of the total number of OTUs on the order of 13,000. We consider this as a relatively low number of OTUs for this globally distributed set of intertidal sediment samples, even more so given that this number might have been somewhat inflated by polymerase chain reaction (PCR) or other methodological errors (Bik *et al.*, 2012; Liu *et al.*, 2017). The generally low OTU number is consistent with the idea that enormous population sizes and unlimited dispersal maintain the global diversity of small eukaryotes to a relatively low level, compared to larger organisms (Fenchel, 2005; Yang *et al.*, 2010).

Although the rarefaction curve for the pooled dataset shows saturation at our observed number of OTUs, individual curves for the four separate regions do not (Fig. 2B). This suggests that approximately 13,000 OTUs are well distributed globally, but that we would have needed deeper sampling or sequencing at each site to observe them all everywhere. Venn diagrams of OTU occurrence also indicate this (Fig. S4 in the Supporting Information). For example, in the whole dataset, 3598 OTUs, or 28% of the total, occurred in all four regions and hence are considered cosmopolitan. A smaller number (2835 of 12,890 OTUs), 22%, were found in only one of the four regions. Within regions, however, the pattern is different. For the four stations in Connecticut, only 14% of OTUs (1264) were common and 37% (3337 of 9104 OTUs) were found only once. In the four stations in Xiamen, 17% of OTUs (1530) were common and 38% (3439 of 9012 OTUs) were unique to one station. It is counterintuitive that a higher percent of OTUs should be cosmopolitan in the global dataset than are in the regional datasets. However, this underscores the sampling problem. In a very large dataset (12 stations with three replicates each), sequencing can be deep enough to saturate, whereas the number of sequences from each regional station is not enough to ensure that all OTUs will be encountered even if they are present. Implied by this is that the abundant taxa, which are more easily encountered even at low sequencing depth, can vary among sampling sites within a region, perhaps suggesting sensitivities of abundance to environmental conditions.

Controlling factors for geographical distributions of rare and abundant microeukaryotes

Our results illustrate that both environmental (deterministic) and spatial (stochastic) processes play significant roles in structuring microeukaryote assemblages in the sandy beach metacommunity. An important ecological aspect of biogeography is to understand the mechanisms that generate and maintain microbial diversity (Hanson *et al.*, 2012). However, unlike the extensive studies of biogeographical patterns in terrestrial ecosystems or other marine

ecosystems, studies on the spatial scales of benthic microeukaryote diversity of sandy beaches are still limited or in their infancy (Fonseca *et al.*, 2014; Gong *et al.*, 2015). Our study provided an opportunity to evaluate potential controlling factors and to determine whether the relative importance of local environment versus spatial factors differ for the rare and abundant sub-communities. We found that the rare sub-community biogeographic patterns were slightly different compared to those for the abundant sub-community because the communities were significantly related to different spatial and environmental factors (Fig. 6). However, variation in both sub-communities was significantly related to temperature, salinity, pH and total nitrogen. Previous biogeographic analyses have shown that climate zone is a main predictor of species richness, increasing from temperate to tropical sandy beaches (Rafael Barboza and Defeo, 2015), but we did not find any direct correlation of different diversity measures with latitude over the approximately 17° of latitude our samples covered. Conversely, in a large-scale meta-analysis salinity was suggested as the major determinant across many (including ocean) ecosystems and to exceed the influence of temperature (Lozupone and Knight, 2007; Liu *et al.*, 2011). We found significant relationships between diversity and both temperature and salinity in our data, though the range in both variables was not great among our sampling sites.

We also observed that both abundant and rare sub-communities had a stronger distance–decay relationship (Fig. 5), although the abundant taxa showed a stronger dispersal ability since more of them were found at multiple sites (Fig. S3 in the Supporting Information). The rare taxa were more diverse than the abundant taxa in sediment samples collected from different regions or even from the same region but different sites. Any dispersal limitation of microeukaryotes should lead to a decrease in community similarity as individuals are more likely to colonize nearby habitats (Lear *et al.*, 2014), leading to the significant distance–decay patterns in this study. Moreover, species sorting that is adaptation to local environments can also lead to distance–decay in community similarity (Hanson *et al.*, 2012), as evidenced by the increased environmental variability with geographic distance (Fig. S8 in the Supporting Information). Hanson and colleagues (2012) highlighted that environmental selection is a major process leading to a distance–decay relationship. We found that the environment plus grain-size factors and spatial factors are nearly equal in explaining the variation of the entire-, abundant- and rare-communities (Fig. 6), but partial Mantel tests showed that both abundant- and rare-communities were mostly governed by environmental and sand (grain size) variables. In addition, grain size had a significant impact on the variation of microeukaryote communities, as evidenced by the explained variation being

slightly lower than that of spatial and other environmental factors (Fig. S7 in the Supporting Information). Several previous studies have shown that species richness significantly increased in response to decreasing sediment particle size, especially in the intertidal zones (Lallias *et al.*, 2015; Rafael Barboza and Defeo, 2015). Grain size and shape can influence both physicochemical properties and biological communities of benthic systems in direct or indirect ways. One abundant group, the ciliates, generally are more abundant in medium and fine sand than in silt and mud, presumably due to the restricted pore size and permeability (Patterson *et al.*, 1989). We found quite a bit of variation in grain size among our sample sites, with QD being the coarsest and ZS being the finest (Fig. S9 in the Supporting Information).

Our analysis revealed that the largest explained fraction of the community variation in both total and abundant microeukaryotes was attributed to the joint effect of spatial and environmental factors. Although we set out to examine primarily spatial factors by focusing on a single habitat, sandy beaches, we did see some differences among sites, and geographic distance was significantly correlated with differences in environmental characteristics (Fig. S8 in the Supporting Information). More importantly, the large-scale distribution patterns in both rare and abundant sub-communities were slightly more influenced by environmental conditions than by dispersal limitation (spatial factor). However, it is notable that the unexplained variation of the rare sub-community (84%) was substantially higher than that of the entire (46%) and abundant (34%) communities. The high proportion of unexplained variation may be due to additional biotic and abiotic factors not measured in this study (Hanson *et al.*, 2012). In fact, tide intensity and range, beach slope and inshore productivity are known to affect spatial patterns of macro-faunal communities in sandy beaches (McLachlan and Dorvlo, 2005). Moreover, based on neutral theory, ecological drift (stochastic processes of birth, death, colonization and extinction) and evolutionary drift (stochastic genetic change) could contribute to purely spatial effects and unexplained variation (Hanson *et al.*, 2012; Zhou *et al.*, 2014). Most likely, additive or synergistic effects of multiple factors are important in regulating the distribution of microeukaryotes in intertidal sediments across different space and time scales (Lei *et al.*, 2014). Thus, an alternative explanation is that random processes play a much larger role in the ecology and evolution of rare microorganism than they do for the better studied abundant taxa (Bonner, 2013). These findings highlight the need to understand what governs the relative balance between stochastic and deterministic processes and what conditions would lead to stochastic processes overwhelming deterministic processes (Wang *et al.*, 2013). With increasingly powerful survey tools and new theoretical models, obviously, extending the sampling sites and

integrating more environmental factors and conducting manipulative experiments across space and time will be helpful to further advance our understanding of the influence and mechanisms of environmental selection and dispersal processes on the biogeography of benthic intertidal microeukaryotes. Additional investigations based on rare taxa are especially needed in order to better understand benthic microeukaryotic diversity and biogeography.

Conclusions

Sampling sandy beach intertidal sediments over a range of spatial scales from 0.001 to 12,000 km, we found DNA sequences representing all of the major eukaryotic clades. Arthropods, diatoms and alveolates (ciliates and dinoflagellates) had the most sequences. At 12,890 OTUs (clustered at 97%), the rarefaction curve for the global dataset showed near saturation and about 30 OTUs were cosmopolitan in all replicate samples. Spatial factors (distance) explained more of the variability among sites for the abundant microeukaryotic OTUs, compared to the rare OTUs, suggesting that many of the rare forms are present at a site because of high dispersal, but do not become abundant under the extant environmental conditions at time of sampling. This was also indicated by the fact that some groups (e.g., fungi and non-metazoan opisthokonts) had high OTU richness represented by relatively few sequences. Significant distance–decay relationships for community similarity were found for both rare and abundant taxa. Our results also provide strong evidence that both environmental (habitat suitability) and spatial (dispersal) factors contributed to the distribution patterns in both rare and abundant sub-communities. However, unlike the well-explained variation of the abundant sub-community, the high proportion of unexplained variation in the rare sub-community suggested more complex mechanisms that may generate and maintain the rare biosphere diversity in the dynamic habitat of sandy beaches.

Experimental procedures

Study sites and sampling design

In early September 2014, 36 benthic samples (12 beaches times three replicates each) were collected from the intertidal sandy zone (Fig. 1 and Table S1 in the Supporting Information). Three replicate samples, each composed of a single core (inner diameter of 6.4 cm, length of 10 cm, volume of about 300 ml), were taken approximately 1 m apart from each other when the sampled area was still just inundated by the falling tide, and the sampled area was approximately half way between the low and high tide. An additional core and pore water sample was taken in each replicate for physicochemical analysis. The samples were transported to the laboratory and processed within 4 h of sampling. For microeukaryotic community analyses, surface (0–5 cm) sediment samples were immersed and shaken in 5 l of sterile seawater over five times

and then pre-filtered with a 200- μm pore size sieve in order to mechanically remove sand grains and large metazoans. Finally, the water sample with microeukaryotes (smaller than 200 μm) was filtered onto a 0.22- μm pore size, 47-mm diameter polycarbonate membrane (Millipore, Billerica, MA, USA). The membranes were stored in a 2 ml tube and kept at -80°C until DNA extraction. To limit DNA contamination, the sterile seawater was pre-filtered through a nominal 0.22 μm pore size membrane (Millipore, Billerica, MA, USA). Filtration equipment was rinsed with sterile water before each sample filtering and cleaned after each filtration with a 10% bleach solution.

Physicochemical analyses

Pore-water temperature, salinity and pH were measured *in situ* with the Hydrolab multi-parameter water quality sondes (HACH, Loveland, CO, USA). The concentrations of TN, total carbon (TC) and total organic carbon (TOC) were determined by a TOC/TN-VCPH analyzer (Shimadzu, Kyoto, Japan). A spectrophotometric method was used to measure TP after digestion. The concentrations of nitrite and nitrate nitrogen ($\text{NO}_x\text{-N}$) and $\text{PO}_4\text{-P}$ were determined with a Lachat QC8500 Flow Injection Autoanalyzer (Lachat Instruments, HACH, Loveland, CO, USA). Sediment properties (sedimentary organic matter content and grain size) were analysed as follows: 10 g of sediment was taken from each replicate sample, the sediment dry weight was determined before and after combustion to a constant weight at 560°C and the fraction lost by combustion was used to estimate the percentage of organic matter. The remaining sediment sample (about 300 g) was used to analyse grain size, by using the dry-sieving method (Bale and Kenny, 2005).

DNA extraction, PCR and high-throughput sequencing

Membranes with microeukaryotes were cut into small pieces, soaked in 1.5 ml of lysis buffer (0.1 M EDTA and 1% SDS) and incubated at room temperature for four hours before DNA extraction. Total DNA was extracted using the FastDNA SPIN Kit and the FastPrep Instrument (MP Biomedicals, Santa Ana, CA, USA) according to the manufacturer's instructions. The universal meta-barcoding primers 1380F and 1510R (Amaral-Zettler *et al.*, 2009) were used to amplify the hyper-variable V9 region of eukaryotic 18S rDNA. Each DNA sample was separately amplified with the PCR, as described by Liu and colleagues (2017). PCR was carried out in 30 μl reactions with 15 μl of Phusion® High-Fidelity PCR Master Mix (New England Biolabs, Beverly, MA, USA), 0.2 μM of forward and reverse primers and about 10 ng template DNA. Thermal cycling consisted of initial denaturation at 98°C for 1 min, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 50°C for 30 s and elongation at 72°C for 60 s. A final elongation step was done at 72°C for 10 min. The PCR products were checked on 2% agarose gel and purified with GeneJET Gel Extraction Kit (Thermo Scientific, Hudson, NH, USA). Sequencing libraries were generated using Next® Ultra™ DNA Library Prep Kit for Illumina (New England Biolabs, Beverly, MA, USA) following manufacturer's recommendations and barcode indexes were added. The library quality was assessed on the Qubit® 2.0 Fluorometer (Thermo Scientific)

and Agilent Bioanalyzer 2100 system (Agilent Technologies, Santa Clara, CA, USA). Finally, the libraries were subjected to Illumina MiSeq sequencing in a paired-end 250-bp sequence read run (Illumina, San Diego, CA, USA).

Bioinformatics

Paired-end reads were assigned to each of the 36 samples according to their unique barcodes. The paired-end reads were merged with FLASH (Magoč and Salzberg, 2011). Raw sequence reads were analysed and quality filtered in MOTHUR v.1.33.3 (Schloss *et al.*, 2009) and QIIME v.1.8.0 (Caporaso *et al.* 2010), with procedures as suggested by Bokulich and colleagues (2013). Sequences of length <135 or >152 , average quality <30 , ambiguous bases >0 or homopolymer length >6 were removed. The remaining high-quality sequence reads were aligned to a reference alignment, and those sequences that did not align to the correct region were eliminated. Chimeras were identified and removed using UCHIME (Edgar *et al.*, 2011). Afterward, a Bayesian classifier was used to classify sequences to the phylum or subphylum level (e.g., Discoba and Ciliophora, respectively) against the SILVA 115 database (Pruesse *et al.*, 2007), using an 80% bootstrap confidence score (Wang *et al.*, 2007). Finally, sequences were split into lower taxonomic ranks using the UPARSE algorithms and assigned to OTUs at a 97% sequence similarity threshold. Singleton OTUs were discarded before the downstream analyses as potential sequencing errors. To enable comparisons between samples, we used a randomly subsampled subset of 34,239 sequences from each sample to standardize sequencing effort across samples. Thus, our final data set retained 1,232,604 sequence reads for the entire metacommunity.

Rare and abundant taxa

Defining the rare biosphere has been arbitrary (Lynch and Neufeld, 2015). In this study, the thresholds for regionally abundant and rare taxa or OTUs were defined based on relative abundance cut-offs (Logares *et al.*, 2014; Liu *et al.*, 2015). OTUs with a mean relative abundance of $>0.1\%$ of the metacommunity were herein defined as abundant taxa, whereas those with a mean relative abundance of $<0.001\%$ are defined as rare taxa. Using these values, our data exactly matched with the previous view (Logares *et al.*, 2014) that abundant taxa contribute $<5\%$ richness and $>50\%$ relative abundance, while rare taxa contribute $>50\%$ richness and $<5\%$ abundance (Table S2 in the Supporting Information).

Analysis of alpha- and beta-diversities

Rarefaction curves and alpha diversity indices of ACE, Chao1, Shannon, Simpson and Pielou's evenness were calculated in the Vegan package of R (R Core Team, 2015), while the Good's coverage was calculated in MOTHUR. The heatmaps showing relative taxonomic abundance were generated in the pheatmap package of R. OTU networks were constructed and graphically edited in Cytoscape (Shannon *et al.*, 2003) using the layout 'edge-weighted spring embedded' with weights.

For comparing communities, non-metric multidimensional scaling ordination (nMDS) analyses were conducted based on Bray–Curtis similarity. Furthermore, an analysis of similarity (ANOSIM) was used to statistically test for significant differences in microeukaryotic communities among the four regions, based on geographical distance. In this analysis, complete separation is indicated by $R = 1$, whereas $R = 0$ suggests no separation. Both nMDS and ANOSIM were performed in PRIMER 7.0 (Clarke and Gorley, 2015).

Relationships between community composition, environmental and spatial factors

Before multivariate statistical analyses, all environment variables, with the exception of pH, were $\log(x + 1)$ transformed to improve normality and homoscedasticity. Spearman's rank correlations were used to determine the relationships between the Bray–Curtis similarity of microeukaryotic communities and the geographical distances among sampling sites and the relationship between the Euclidean distances among environmental variables and the geographical distances.

A set of spatial variables was generated using PCNM analysis (Borcard and Legendre, 2002) based on the longitude and latitude of sampling sites. RDA or CCA was performed to explore the relationships between microeukaryotic communities and spatial (based on PCNM variables), environmental and grain size variables, based on the longest gradient lengths of DCA. Before the RDA or CCA analysis, forward selection procedures were used to select subsets of spatial, environmental and grain size variables, respectively, by using the 'ordiR2step' function from the vegan package in R (Blanchet *et al.*, 2008). All non-significant ($P > 0.05$) variables were eliminated in downstream analyses. The relative importance of the spatial, environmental and grain size factors and their co-variations, for explaining differences in community composition was distinguished using variation partitioning with adjusted F^2 coefficients (Wang *et al.*, 2015; Peres-Neto *et al.*, 2006). In this analysis, the residual fraction equals the unexplained variance.

To compare the variance partitioning results, we also conducted both standard and partial Mantel tests for relationship between microeukaryotic community similarity and spatial, environmental and grain size variables (Legendre and Legendre, 2012). All statistics analyses were performed in R (<http://www.r-project.org>) unless otherwise indicated (R Core Team, 2015).

In our study, the environmental factors (e.g., temperature) are continuous variables, whereas the grain size (which also is an environmental factor) is a categorical variable and for this reason we analysed it separately from the other factors.

Accession number

All raw sequence data from this study have been deposited in the GenBank's Sequence Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/>) under the accession number SRP099521 (BioProject Accession PRJNA362750 and Bio-Sample accession SAMN06246921–SAMN06246956).

Acknowledgements

We thank Luke Sanchez, Luciana Santoferrara and Ewelina Rubin from the Department of Marine Sciences, University of Connecticut and Sheng-Fang Tsai from Institute of Marine Environmental Chemistry and Ecology, National Taiwan Ocean University for experimental help. We also thank David Wilkinson from the School of Life Sciences, University of Lincoln, UK, for comments on the manuscript. This study was supported by the National Natural Science Foundation of China (41276133), the Natural Science Foundation of Fujian Province (2014J01163) and the Fundamental Research Funds for the Central Universities (20720160115). Additional support was provided by the University of Connecticut and by a grant to GBM from the US National Science Foundation (OCE14735515). The authors declare no competing commercial interests in relation to the submitted work.

References

- Adl, S.M., Simpson, A.G., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., *et al.* (2012) The revised classification of eukaryotes. *J Eukaryot Microbiol* **59**: 429–514.
- Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One* **4**: e6372.
- Appeltans, W., Ah Yong Shane, T., Anderson, G., Angel Martin, V., Artois, T., Bailly, N., *et al.* (2012) The magnitude of global marine species diversity. *Curr Biol* **22**: 2189–2202.
- Bale, A.J., and Kenny, A.J. (2005) Sediment analysis and seabed characterization. In *Methods for the Study of Marine Benthos*. Eleftheriou, A., and McIntyre, A. (eds). Oxford, UK: Blackwell Science Ltd, pp. 43–81.
- Bates, S.T., Clemente, J.C., Flores, G.E., Walters, W.A., Parfrey, L.W., Knight, R., *et al.* (2013) Global biogeography of highly diverse protistan communities in soil. *ISME J* **7**: 652–659.
- Bik, H.M., Porazinska, D.L., Creer, S., Caporaso, J.G., Knight, R., and Thomas, W.K. (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol Evol* **27**: 233–243.
- Blanchet, F.G., Legendre, P., and Borcard, D. (2008) Forward selection of explanatory variables. *Ecology* **89**: 2623–2632.
- Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., *et al.* (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* **10**: 57–59.
- Bonner, J.T. (2013) *Randomness in evolution*. Princeton, USA: Princeton University Press.
- Borcard, D., and Legendre, P. (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Model* **153**: 51–68.
- del Campo, J., Mallo, D., Massana, R., Vargas, C., Richards, T.A., and Ruiz-Trillo, I. (2015) Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. *Environ Microbiol* **17**: 3195–3207.
- 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.
- Cermeno, P., and Falkowski, P.G. (2009) Controls on diatom biogeography in the ocean. *Science* **325**: 1539–1541.
- Clarke, K.R., and Gorley, R.N. (2015) *PRIMER v7: User Manual/Tutorial*. Plymouth, UK: PRIMER-E Ltd.
- Decelle, J., Romac, S., Sasaki, E., Not, F., and Mahe, F. (2014) Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (Radiolarians) assessed by high-throughput sequencing. *PLoS One* **9**: e104297.
- Defeo, O., and McLachlan, A. (2005) Patterns, processes and regulatory mechanisms in sandy beach macrofauna: a multi-scale analysis. *Mar Ecol Prog Ser* **295**: 1–20.
- Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.
- Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., et al. (2011a) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J* **5**: 1344–1356.
- Edgcomb, V.P., Beaudoin, D., Gast, R., Biddle, J.F., and Teske, A. (2011b) Marine subsurface eukaryotes: the fungal majority. *Environ Microbiol* **13**: 172–183.
- Fenchel, T. (2005) Where are all the species? *Environ Microbiol* **7**: 473–475.
- Fonseca, V.G., Carvalho, G.R., Sung, W., Johnson, H.F., Power, D.M., Neill, S.P., et al. (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nat Commun* **1**: 98.
- Fonseca, V.G., Carvalho, G.R., Nichols, B., Quince, C., Johnson, H.F., Neill, S.P., et al. (2014) Metagenetic analysis of patterns of distribution and diversity of marine meiobenthic eukaryotes. *Global Ecol Biogeogr* **23**: 1293–1302.
- Gobet, A., Böer, S.I., Huse, S.M., van Beusekom, J.E., Quince, C., Sogin, M.L., et al. (2012) Diversity and dynamics of rare and of resident bacterial populations in coastal sands. *ISME J* **6**: 542–553.
- Gong, J., Shi, F., Ma, B., Dong, J., Pachiadaki, M., Zhang, X., et al. (2015) Depth shapes α - and β -diversities of microbial eukaryotes in surficial sediments of coastal ecosystems. *Environ Microbiol* **17**: 3722–3737.
- Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
- Harder, C.B., Rønn, R., Brejnrod, A., Bass, D., Al-Soud, W.A., and Ekelund, F. (2016) Local diversity of heathland Cercozoa explored by in-depth sequencing. *ISME J* **10**: 2488–2497.
- Harris, L., Nel, R., Holness, S., Sink, K., and Schoeman, D. (2014) Setting conservation targets for sandy beach ecosystems. *Estuar Coast Shelf Sci* **150**: 45–57.
- Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Dufournel, I.J., 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 USA* **110**: 6004–6009.
- Lallias, D., Hiddink, J.G., Fonseca, V.G., Gaspar, J.M., Sung, W., Neill, S.P., et al. (2015) Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *ISME J* **9**: 1208–1221.
- Lear, G., Bellamy, J., Case, B.S., Lee, J.E., and Buckley, H.L. (2014) Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. *ISME J* **8**: 1715–1726.
- Legendre, P., and Legendre, L. (2012) *Numerical Ecology*, 3rd edn. Amsterdam, The Netherlands: Elsevier.
- Lei, Y.L., Stumm, K., Wickham, S.A., and Berninger, U.G. (2014) Distributions and biomass of benthic ciliates, foraminifera and amoeboid protists in marine, brackish, and freshwater sediments. *J Eukaryot Microbiol* **61**: 493–508.
- Liu, L.M., Yang, J., and Zhang, Y.Y. (2011) Genetic diversity patterns of microbial communities of in a subtropical riverine ecosystem (Jiulong River, southeast China). *Hydrobiologia* **678**: 113–125.
- Liu, L.M., Yang, J., Yu, Z., and Wilkinson, D.M. (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J* **9**: 2068–2077.
- Liu, L.M., Liu, M., Wilkinson, D.M., Chen, H.H., Yu, X.Q., and Yang, J. (2017) DNA metabarcoding reveals that 200 μ m size-fractionated filtering is unable to discriminate between planktonic microbial and large eukaryotes. *Mol Ecol Resour* **17**: 991–1002.
- 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.
- Lozupone, C.A., and Knight, R. (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci USA* **104**: 11436–11440.
- Lynch, M.D., and Neufeld, J.D. (2015) Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* **13**: 217–229.
- Magoč, T., and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957–2963.
- Margules, C.R., and Pressey, R.L. (2000) Systematic conservation planning. *Nature* **405**: 243–253.
- Márquez, L.M., Miller, D.J., MacKenzie, J.B., and van Oppen, M.J. (2003) Pseudogenes contribute to the extreme diversity of nuclear ribosomal DNA in the hard coral *Acropora*. *Mol Biol Evol* **20**: 1077–1086.
- 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.
- Massana, R., and Pedrós-Alió, C. (2008) Unveiling new microbial eukaryotes in the surface ocean. *Curr Opin Microbiol* **11**: 213–218.
- Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., et al. (2015) Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ Microbiol* **17**: 4035–4049.
- McArdle, S.B., and McLachlan, A. (1992) Sand beach ecology: swash features relevant to the macrofauna. *J Coast Res* **8**: 398–407.
- McLachlan, A., and Brown, A.C. (2006) *The Ecology of Sandy Shores*. Amsterdam, The Netherlands: Elsevier.
- McLachlan, A., and Dorvlo, A. (2005) Global patterns in sandy beach macrobenthic communities. *J Coast Res* **21**: 674–687.
- Nel, R., Campbell, E.E., Harris, L., Hauser, L., Schoeman, D.S., McLachlan, A., et al. (2014) The status of sandy

- beach science: past trends, progress, and possible futures. *Estuar Coast Shelf Sci* **150**: 1–10.
- Nunes-Alves, C. (2015) Deep sequencing of the global oceans. *Nat Rev Genet* **16**: 378.
- Orsi, W.D., Edgcomb, V.P., Christman, G.D., and Biddle, J.F. (2013) Gene expression in the deep biosphere. *Nature* **499**: 205–208.
- Pan, Y., Wang, L., Zhang, W., Liu, G., and Lin, S. (2016) Genetic analysis of *Noctiluca scintillans* populations indicates low latitudinal differentiation in China but high China–America differences. *J Exp Mar Biol Ecol* **477**: 31–39.
- Patterson, D., Larsen, J., and Corliss, J. (1989) The ecology of heterotrophic flagellates and ciliates living in marine sediments. *Prog Protistol* **3**: 185–277.
- Pedrós-Alió, C. (2012) The rare bacterial biosphere. *Annu Rev Mar Sci* **4**: 449–466.
- Peres-Neto, P.R., Legendre, P., Dray, S., and Borcard, D. (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**: 2614–2625.
- Pernice, M.C., Logares, R., Guillou, L., and Massana, R. (2013) General patterns of diversity in major marine microeukaryote lineages. *PLoS One* **8**: e57170.
- Pernice, M.C., Giner, C.R., Logares, R., Perera-Bel, J., Acinas, S.G., Duarte, C.M., *et al.* (2015) Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME J* **10**: 945–958.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.
- R Core Team. (2015) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL <http://www.R-project.org/>.
- Rafael Barboza, F., and Defeo, O. (2015) Global diversity patterns in sandy beach macrofauna: a biogeographic analysis. *Sci Rep* **5**: 14515.
- Richards, T.A., Jones, M.D., Leonard, G., and Bass, D. (2012) Marine fungi: their ecology and molecular diversity. *Annu Rev Mar Sci* **4**: 495–522.
- Scheckenbach, F., Hausmann, K., Wylezich, C., Weitere, M., and Arndt, H. (2010) Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. *Proc Natl Acad Sci USA* **107**: 115–120.
- 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.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504.
- Short, A.D. (1996) The role of wave height, period, slope, tide range and embaymentisation in beach classifications: a review. *Rev Chil Hist Nat* **69**: 589–604.
- Simon, M., López-García, P., Deschamps, P., Moreira, D., Restoux, G., Bertolino, P., *et al.* (2015) Marked seasonality and high spatial variability of protist communities in shallow freshwater systems. *ISME J* **9**: 1941–1953.
- Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al.* (2006) Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci USA* **103**: 12115–12120.
- Telford, R.J., Vandvik, V., and Birks, H.J.B. (2006) Dispersal limitations matter for microbial morphospecies. *Science* **312**: 1015.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., *et al.* (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605.
- Wang, J., Shen, J., Wu, Y., Tu, C., Soininen, J., Stegen, J.C., *et al.* (2013) Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *ISME J* **7**: 1310–1321.
- Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.
- Wang, Y.M., Yang, J., Liu, L.M., and Yu, Z. (2015) Quantifying the effects of geographical and environmental factors on distribution of stream bacterioplankton within nature reserves of Fujian, China. *Environ Sci Pollut Res* **22**: 11010–11021.
- Worden, A.Z., Follows, M.J., Giovannoni, S.J., Wilken, S., Zimmerman, A.E., and Keeling, P.J. (2015) Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**: 1257594.
- Worm, B., Barbier, E.B., Beaumont, N., Duffy, J.E., Folke, C., Halpern, B.S., *et al.* (2006) Impacts of biodiversity loss on ocean ecosystem services. *Science* **314**: 787–790.
- Yang, J., Smith, H.G., Sherratt, T.N., and Wilkinson, D.M. (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.
- Yu, L.Y., Zhang, W.J., Liu, L.M., and Yang, J. (2015) Determining microeukaryotic plankton community around Xiamen Island, southeast China, using Illumina MiSeq and PCR-DGGE techniques. *PLoS One* **10**: e0127721.
- Zhang, Q., He, X., and Yan, T. (2015) Differential decay of wastewater bacteria and change of microbial communities in beach sand and seawater microcosms. *Environ Sci Technol* **49**: 8531–8540.
- Zhou, J., Deng, Y., Zhang, P., Xue, K., Liang, Y., Van Nostrand, J.D., *et al.* (2014) Stochasticity, succession, and environmental perturbations in a fluidic ecosystem. *Proc Natl Acad Sci USA* **111**: E836–E845.
- Zubkov, M.V., and Tarran, G.A. (2008) High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature* **455**: 224–226.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's website.

Fig. S1. Latitudinal trends in intertidal benthic microeukaryotic diversity (OTU number, ACE, Chao1, Shannon, Simpson, and Pielou evenness). One-way ANOVA was used to compare the differences among four regions, error bars are standard error and *P* values are given. Subsequently, ANOVA was used among 12 stations (three replicates in each station) and all *P* values are higher than 0.05.

Fig. S2. Frequency distributions of all, abundant and rare OTUs from 36 samples. (A) All OTUs. Networks representing regionally abundant (B) and rare (C) taxa, respectively. Nodes represent OTUs that may connect different samples through edges (lines).

Fig. S3. Spearman's rank correlation between medians of microeukaryotic OTUs abundance and number of sites occupied (n is the number of OTUs).

Fig. S4. Venn diagram showing the unique and shared OTUs in the regional and local communities. The operational taxonomic units (OTUs) were defined at 97% sequence similarity threshold.

Fig. S5. Spearman's rank correlations between the Bray–Curtis similarity of microeukaryotic community and geographical distance between sampling sites (n is the number of comparison and P values are indicated), [QD], [ZS] and [XM] regions located at China and [CT] region located at USA. All, all microeukaryotic taxa; abundant, abundant taxa; rare, rare taxa.

Fig. S6. RDA/CCA ordinations showing the microeukaryotic community composition in relation to significant spatial, environmental and sand size variables ($P < 0.05$). TN, total nitrogen; TP, total phosphorus; PO₄-P, phosphate phosphorus; ORG, organic matter content; S1–S6, different sand

sizes (S1, S2, S3, S4, S5 and S6 indicate >2 , 1–2, 0.5–1, 0.25–0.5, 0.1–0.25 and 0.075–0.1 mm, respectively). All: all microeukaryotic taxa; abundant: abundant taxa; rare: rare taxa.

Fig. S7. Variation partitioning among spatial, environmental and sand grain size variables for the three sets of data (All: all microeukaryotic taxa community; abundant: abundant sub-community; rare: rare sub-community). Forward selection procedures were used to select the best subset of spatial, environmental and sand size variables explaining community variation, respectively.

Fig. S8. Spearman's rank correlation between the Euclidean distance and geographical distance (n is the number of comparison, all 10 environmental variables were used for Euclidean distance, see 'Experimental procedures' section).

Fig. S9. Sand size fraction distribution based on weight percentage for the 36 sediment samples.

Table S1. Abundant and rare OTUs information for the 36 benthic microeukaryotic communities based on the Illumina dataset.

Table S2. General descriptions of all, abundant and rare OTUs data sets at 97% sequence similarity level.

Table S3. Diversity, richness and coverage of benthic microeukaryotic communities in each of the individual samples.