Current Biology 24, 813–821, April 14, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.02.050

Article

Patterns of Rare and Abundant Marine Microbial Eukaryotes

Ramiro Logares,[1,*] Stéphane Audic,[2,3] David Bass,[4] Lucie Bittner,[2,3,5] Christophe Boutte,[2,3] Richard Christen,[6,7] Jean-Michel Claverie,[8] Johan Decelle,[2,3] John R. Dolan,[9] Micah Dunthorn,[5] Bente Edvardsen,[10] Angélique Gobet,[2,3] Wiebe H.C.F. Kooistra,[11] Frédéric Mahé,[2,3,5] Fabrice Not,[2,3] Hiroyuki Ogata,[8,12] Jan Pawlowski,[13] Massimo C. Pernice,[1] Sarah Romac,[2,3] Kamran Shalchian-Tabrizi,[10] Nathalie Simon,[2,3] Thorsten Stoeck,[5] Sébastien Santini,[8] Raffaele Siano,[14] Patrick Wincker,[15] Adriana Zingone,[11] Thomas A. Richards,[16] Colomban de Vargas,[2,3] and Ramon Massana[1]

[1]Institut de Ciències del Mar (ICM), CSIC, Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain
[2]ADMM UMR 7144, UPMC Paris 06, Station Biologique de Roscoff, 29682 Roscoff, France
[3]ADMM UMR 7144, CNRS, Station Biologique de Roscoff, 29682 Roscoff, France
[4]Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK
[5]Department of Ecology, University of Kaiserslautern, 67663 Kaiserslautern, Germany
[6]SAE UMR 7138, CNRS, Parc Valrose BP71, 06108 Nice Cedex 02, France
[7]SAE UMR 7138, Université de Nice-Sophia Antipolis, Parc Valrose BP71, 06108 Nice Cedex 02, France
[8]IGS UMR 7256, CNRS, Aix-Marseille Université, 13288 Marseille, France
[9]LOV UMR 7093, CNRS, UPMC Paris 06, 06230 Villefranche-sur-Mer, France
[10]Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway
[11]Ecology and Evolution of Plankton, Stazione Zoologica Anton Dohrn, Villa Comunale 1, 80121 Naples, Italy
[12]Education Academy of Computational Life Sciences, Tokyo Institute of Technology, Tokyo 152-8552, Japan
[13]Department of Genetics and Evolution, University of Geneva, 1211 Geneva, Switzerland
[14]Ifremer, Centre de Brest, DYNECO/Pelagos BP70, 29280 Plouzané, France
[15]CEA, Genoscope, 2 Rue Gaston Crémieux, 91000 Evry, France
[16]Biosciences, University of Exeter, Geoffrey Pope Building, Exeter EX4 4QD, UK

Summary

Background: Biological communities are normally composed of a few abundant and many rare species. This pattern is particularly prominent in microbial communities, in which most constituent taxa are usually extremely rare. Although abundant and rare subcommunities may present intrinsic characteristics that could be crucial for understanding community dynamics and ecosystem functioning, microbiologists normally do not differentiate between them. Here, we investigate abundant and rare subcommunities of marine microbial eukaryotes, a crucial group of organisms that remains among the least-explored biodiversity components of the biosphere. We surveyed surface waters of six separate coastal locations in Europe, independently considering the picoplankton, nanoplankton, and microplankton/mesoplankton organismal size fractions.

Results: Deep Illumina sequencing of the 18S rRNA indicated that the abundant regional community was mostly structured by organismal size fraction, whereas the rare regional community was mainly structured by geographic origin. However, some abundant and rare taxa presented similar biogeography, pointing to spatiotemporal structure in the rare microeukaryote biosphere. Abundant and rare subcommunities presented regular proportions across samples, indicating similar species-abundance distributions despite taxonomic compositional variation. Several taxa were abundant in one location and rare in other locations, suggesting large oscillations in abundance. The substantial amount of metabolically active lineages found in the rare biosphere suggests that this subcommunity constitutes a diversity reservoir that can respond rapidly to environmental change.

Conclusions: We propose that marine planktonic microeukaryote assemblages incorporate dynamic and metabolically active abundant and rare subcommunities, with contrasting structuring patterns but fairly regular proportions, across space and time.

Introduction

Microbes are the dominant form of life in the oceans, playing fundamental roles in ecosystem functioning and biogeochemical processes on local and global scales [1–4]. However, limited knowledge of their diversity and community structure across space and time [5, 6] hinders our understanding of the links between microbial life and ecosystem functioning [7]. During the last decade, technological progress in molecular ecology and environmental sequencing has substantially boosted our understanding of marine microbes, unveiling notable patterns of abundant and rare subcommunities [4, 8, 9], reminiscent of patterns observed in classical plant and animal ecology [10]. The recently discovered large amount of rare taxa in microbial communities is now referred to as the "rare biosphere" [11], and its exploration is made feasible today by means of high-throughput sequencing (HTS) technologies [12].

Abundant and rare microbial subcommunities may have fundamentally different characteristics and ecological roles. For example, rare marine microbes are hypothesized to include ecologically redundant taxa that could increase in abundance following environmental perturbation or change and maintain continuous ecosystem functioning [13]. Locally rare taxa can also act as seeds for seasonal succession or sporadic blooms. Conversely, the drastic decrease in abundance or even extinction of a globally abundant oceanic microbe with no ecologically comparable counterpart in the rare biosphere could have significant and unpredictable effects on the global ecosystem. Most of the studies to date

*Correspondence: ramiro.logares@gmail.com

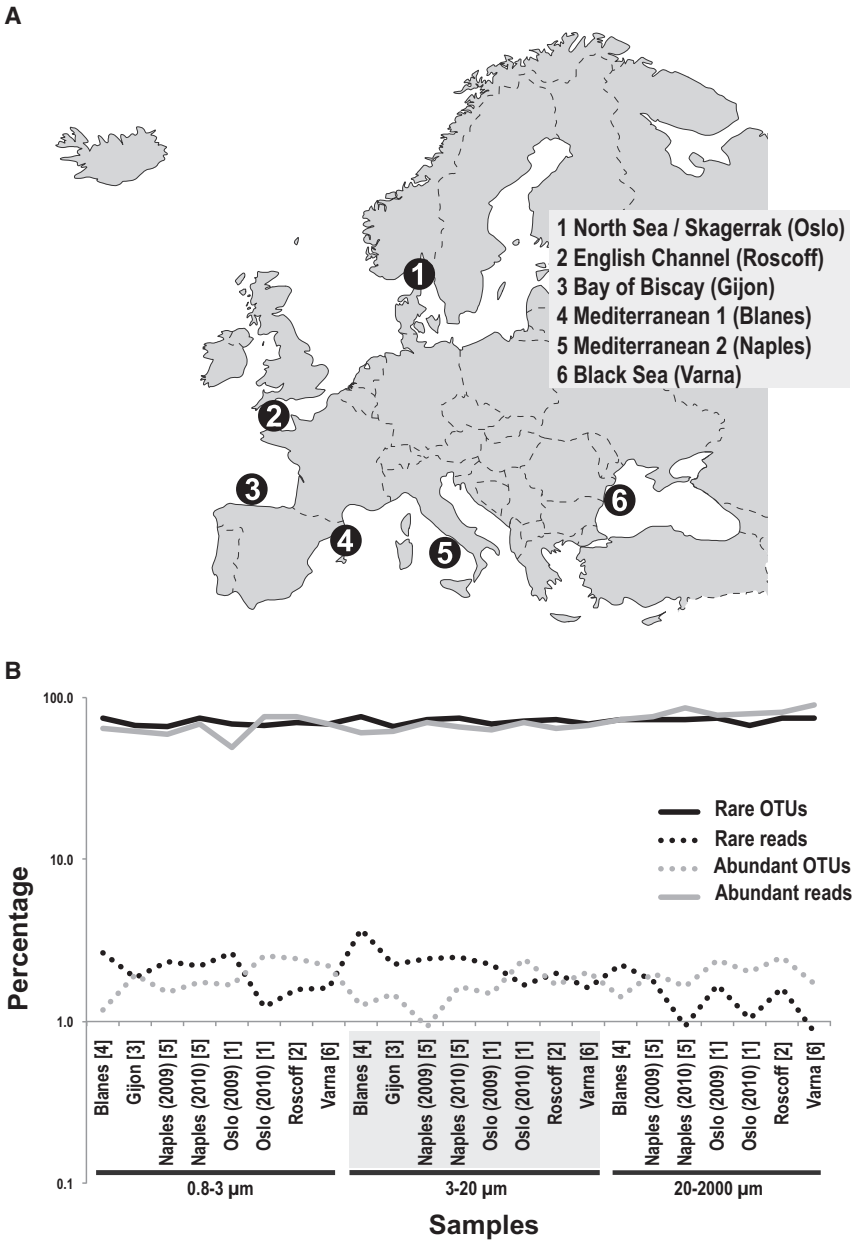


Figure 1. Communities Displayed Regularity in the Proportions of Locally Abundant and Rare OTUs

(A) Sampled locations from the North Sea to the Black Sea.

(B) Percentage of locally abundant (>1%) and rare (<0.01%) OTUs and the corresponding Illumina reads across all samples, indicating different organismal size fractions (in μm) and geographic locations (in brackets) according to (A).

See also Tables S1, S2, and S3; Figures S1, S2, and S4.

that have differentiated between abundant and rare marine microbial subcommunities concern bacteria. Abundant bacteria contribute mostly to biomass, carbon flow, and nutrient cycling, whereas the generally large numbers of rare bacteria contribute predominantly to species richness [9]. Different strategies are observed among rare bacteria, such as dormant or inactive taxa that grow exponentially when the right conditions are met [8, 14–16] or taxa that seem to remain members of the rare biosphere [15, 17] even when they have high relative metabolic activity. Rare bacteria may perform crucial ecosystem functions [18], and some of them can be metabolically more active than abundant taxa in the same community [14, 15]. Both abundant and rare bacteria can present similar biogeographic patterns [17, 19], indicating similar community assembly mechanisms.

Compared to bacteria, we know even less about abundant and rare marine microbial eukaryote (protist) subcommunities, and, overall, marine protists remain one of the least-explored features of natural biodiversity [20]. Recent studies using 454 pyrosequencing [12] recovered few dominant protist taxa and a large number of rare ones from specific marine and freshwater communities [21–25]. Little information is available regarding protist metabolic activity. A recent study comparing rRNA/rDNA ratios (a proxy of microbial activity) in freshwater lakes suggested that, in contrast to bacteria, dormancy does not play an important role in planktonic protist communities [14].

Here, we explore fundamental patterns of rare and abundant marine planktonic protistan subcommunities occurring along the European coastline, from the North Sea (Norway) to the Black Sea (Bulgaria) (Figure 1A). Using Illumina [12] and, to a more limited extent, 454 HTS platforms, we generated a large data set of both 18S rRNA and rDNA tags based on total RNA and DNA extracts from three organismal size fractions: the picoplankton (0.8–3 μm), nanoplankton (3–20 μm), and microplankton/mesoplankton (20–2,000 μm) [3, 26]. The wide geographical and organismal scales of this data set, combined with ultra-deep sequencing, allowed us to address the following main questions. Are the relative proportions of abundant and rare protist subcommunities fluctuating across space and time? What structural and biogeographic patterns are observed in these subcommunities? Do locally abundant taxa tend to be regionally abundant? Are there specific phylogenetic and activity patterns associated with abundant and/or rare marine protistan subcommunities? We found that abundant and rare assemblages present contrasting structuring patterns and phylogenetic characteristics, despite a remarkable consistency in their relative proportions across individual samples. Furthermore, rare subcommunities included a large number of predominantly active lineages that presented biogeography.

Results

General Patterns of Richness and Evenness

Unless stated otherwise, our results are derived from the Illumina V9 18S rRNA tag data set clustered into 95% similarity

Table 1. General Description of Both Complete and Normalized V9 18S rRNA Illumina Data Sets

	Combined Size Fractions		0.8–3 (μm)		3–20 (μm)		20–2,000 (μm)	
	Nonnormalized Data Set[a]	Normalized Data Set[b]	Nonnormalized Data Set[a]	Normalized Data Set[b]	Nonnormalized Data Set[a]	Normalized Data Set[b]	Nonnormalized Data Set[a]	Normalized Data Set[b]
Number of clean reads	5,696,049	1,794,000	2,279,669	624,000	2,298,280	624,000	1,118,100	546,000
Samples	23	23	8	8	8	8	7	7
Geographic sites	6	6	6	6	6	6	5	5
All OTUs[c]	9,007	7,035	6,597	4,412	7,157	4,786	3,491	2,941
Abundant OTUs[d]	155 (1.7%)	154 (2.2%)	153 (2.3%)	153 (3.5%)	143 (1.9%)	144 (3.0%)	95 (2.7%)	95 (3.2%)
Rare OTUs[e]	7,333 (81%)	5,329 (75.7%)	5,145 (77.9%)	2,981 (67.5%)	5,614 (78.4%)	3,242 (67.7%)	2,432 (69.6%)	1,865 (63.4%)

See also Figures S1, S2, S3, and S5.
[a]Variable number of reads per sample.
[b]78,000 reads per sample in all samples.
[c]All OTUs included in the data set.
[d]OTUs abundant in the regional community; average relative abundances >0.1%.
[e]OTUs rare in the regional community; average relative abundances <0.001%.

operational taxonomic units (OTUs; Table S1 available online) and prepared using RNA extracts. The 95% threshold was selected for all downstream analyses in order to minimize any inflation of diversity estimates [27] caused by remaining tags (if any) with misincorporated nucleotides. In the local community, we defined OTUs as "abundant" when they reached relative abundances above 1% of the tags and "rare" when their abundances were below 0.01%, following other studies in bacteria [9, 17] and protists [25]. In the regional community (combination of local communities), the thresholds for abundant or rare OTUs were >0.1% and <0.001%, respectively. In addition, we tested a 97% OTU clustering threshold, and comparable patterns regarding proportions of locally abundant and rare subcommunities were obtained (Table S2).

In total, ~72% of all OTUs were found only in a single size fraction: the picoplankton, nanoplankton, or microplankton/mesoplankton (Figure S1). Similarly, ~75% of the rare OTUs and ~62% of the abundant OTUs in the regional community were restricted to a single organismal size fraction. This indicates that our seawater filtering protocol, which was used to separate total plankton communities into three distinct organismal size fractions, was effective.

In rarefaction analyses, considering all reads (5.69×10^6) and samples from the regional community, richness (based on 95% similarity OTUs) approached saturation at ~9,000 OTUs (Figure S2). OTU richness also approached saturation in most local communities (800–3,000 OTUs; Figure S2A). The highest richness was observed in the nanoplankton (4,786 OTUs after normalization), and the lowest richness was observed in the microplankton/mesoplankton (2,941 OTUs; Table 1). Evenness was low in the regional community, within different size fractions, and in all studied local communities (Figures S3A, S3B, and S4), with the majority of OTUs being rare and only a few being abundant. In the regional community, considering both pooled and separate size fractions, abundant taxa made up <3.5% of the total OTUs, whereas rare taxa made up >63.4% of the OTUs (Table 1). When considering pooled normalized size fractions, the percentage of total reads falling into rare OTUs in the regional community was 1.1% (20,000 reads), whereas the percentage of total reads falling into abundant OTUs was 80.7% (1,448,079 reads).

Overall, a total of 20 out of 23 analyzed samples fitted the log-normal model [10, 28] of species abundance distribution (SAD), according to the Akaike's information criterion [29] (Figure S4). Assuming a log-normal distribution, we fitted our regional community data to the truncated Preston log-normal model [10, 30] (Figure S5A) and estimated that we recovered 64%–67% of the OTUs in the European coastal region (Figure S5B). Therefore, even though our deep Illumina sequencing approach recovered the majority of OTUs from our sample set, extra sampling effort is needed to recover the total richness of the studied area.

Community Structure across Space, Time, and Organismal Size Fractions

The proportions of locally rare (<0.01%) and abundant (>1%) OTUs (by our definition) were relatively constant across communities (Figure 1B; Table S1), with ranges of 66.2%–76.6% for rare OTUs and 0.9%–2.7% for abundant OTUs (Figure 1B; Table S1). Reads corresponding to locally abundant OTUs represented on average 70.1% (SD = 9.5) of the data set, whereas reads corresponding to rare OTUs represented on average 1.9% (SD = 0.7) (Figure 1B).

β diversity (as described by Bray-Curtis dissimilarities between samples) within rare and abundant regional communities (i.e., pooled rare or abundant subcommunities) showed a moderate but significant correlation when considering normalized OTUs from all size fractions together (Mantel test: $r_{(\text{abundant}|\text{rare})} = 0.73$; $p < 0.001$) and within the picoplankton (Mantel test: $r_{(\text{abundant}|\text{rare, 0.8–3})} = 0.69$; $p < 0.05$). The correlation was weaker, but still significant, in the nanoplankton and microplankton/mesoplankton (Mantel test: $r_{(\text{abundant}|\text{rare, 3–20})} = 0.44$; $r_{(\text{abundant}|\text{rare, 20–2,000})} = 0.46$; $p < 0.05$). These correlations indicate that some abundant and rare taxa share similar biogeography. However, the abundant regional community was structured mostly by size fraction because the OTU composition of abundant microplankton/mesoplankton was more similar among samples of this fraction than to any sample of the picoplankton and nanoplankton (Figure 2A). In contrast, the rare regional community was mostly structured by sampling site, with samples from different organismal size fractions but from the same site being normally more similar in OTU composition when compared to samples from other sites (Figure 2A). Network analyses provided further insight into these patterns by showing that within the abundant regional community, the smaller size fractions (picoplankton

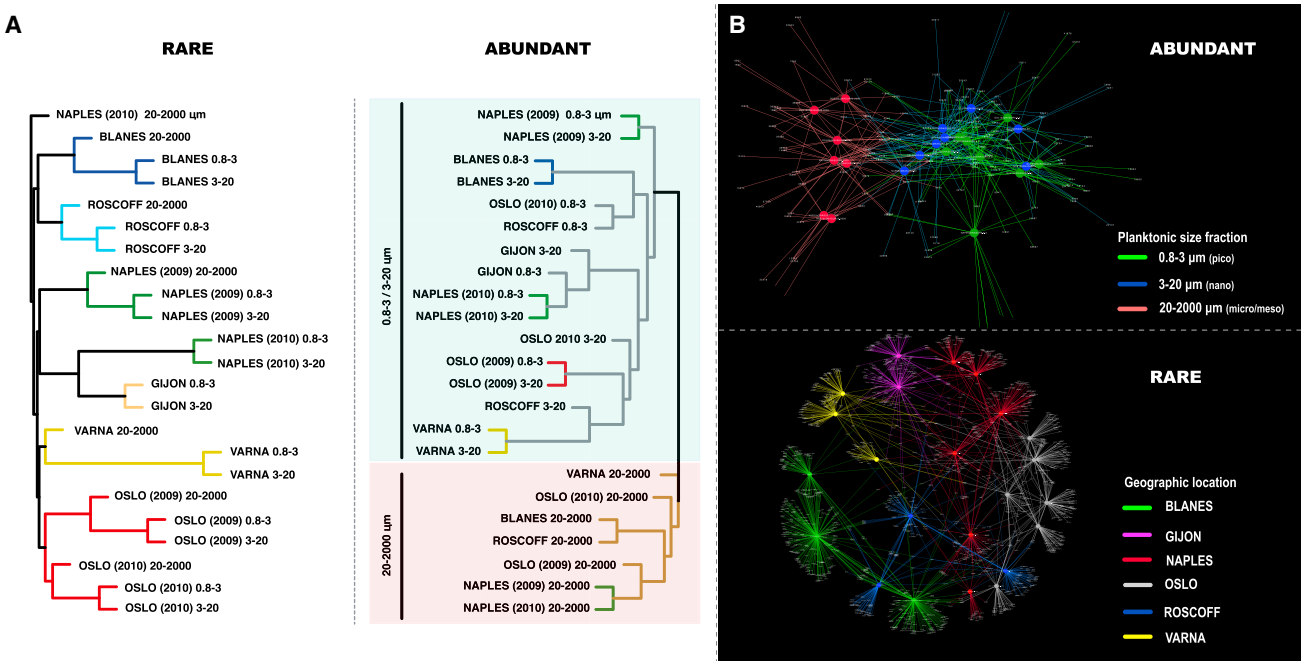


Figure 2. Contrasting Structuring Patterns in Abundant and Rare Regional Communities

(A) Unweighted pair group method with arithmetic mean (UPGMA) dendrograms based on Bray-Curtis dissimilarities between samples (normalized data set) for both the rare and abundant regional communities. Branch colors indicate groups of samples originating from the same geographic site. Note that two large clusters are present within the abundant regional community, separating the picoplankton and nanoplankton from the microplankton/mesoplankton. (B) Networks representing abundant and rare regional communities. Larger nodes (circles) represent samples, whereas smaller nodes represent OTUs that may connect (i.e., may be present in) different samples through edges (lines). The most relevant structuring features for both the abundant and rare regional communities were mapped onto the networks with colors. These structuring features were organismal size fractions (abundant) and geographic origins (rare).

and nanoplankton) shared more OTUs between them than in the larger size fraction (microplankton/mesoplankton; Figure 2B). In contrast, the rare regional community network showed that several OTUs were unique to single samples and that the few shared OTUs tended to be shared among samples of the same site (Figure 2B).

Further exploration across samples of OTUs that were locally abundant in at least one sample (total 175 OTUs) showed that none of them presented abundances >1% in all samples. In analyses of individual size fractions, most OTUs with abundances >1% were abundant at a single site/sample (Figures 3A–3C) and were often rare or of intermediate abundance elsewhere. Only one OTU within the fraction 3–20 µm displayed abundances >1% in all samples (Figure 3B).

Phylogenetic Patterning of Abundant versus Rare Regional Communities

The constructed phylogeny contained 11 reference OTUs that exclusively represented regionally abundant OTUs, 107 reference OTUs that represented both regionally abundant and rare OTUs, and 1,225 reference OTUs that exclusively represented regionally rare OTUs (Figure 4). Whereas the majority of the regionally abundant OTUs had relatively close evolutionary relatives among the rare OTUs, the majority of the regionally rare OTUs had no close evolutionary relatives among the abundant OTUs. Comparisons (using BLAST) of Illumina representative reads from abundant and rare OTUs against each other supported this pattern. About 90% of the abundant OTUs (n = 154) produced significant BLAST hits against the rare OTUs (i.e., hits with coverage >97% and identity >70%), whereas only about 31% of the rare OTUs (n = 5,329) produced significant hits against the abundant OTUs. Faith's phylogenetic diversity (PD) measure [31] was higher in the rare regional community when compared with the abundant community at a similar sampling depth (Figure 4B). Both the mean phylogenetic distance (MPD) and the mean nearest taxon distance (MNTD) [32] indicated that regionally abundant OTUs included in the phylogeny (n = 118) clustered together at a higher frequency than what was expected by chance (Figure 4C). Such a pattern is expected to occur when the environment selects related taxa that share favorable traits [32, 33]. Conversely, the MPD and MNTD among regionally rare OTUs (n = 1,332) did not present deviations from a random distribution (Figure 4C).

Activity versus Abundance

In order to check to what extent the community and phylogenetic patterns described above are due to the use of RNA tags and not DNA tags, we analyzed 15 samples for which both DNA- and RNA-based tags (V4 18S, 454 tags) were obtained. The relative abundance of OTUs in the regional community that were present in both the DNA and RNA data sets showed on average a nearly 1:1 relationship (Figure 5A). Both the DNA and RNA recovered a number of OTUs that were consistently rare or abundant in the regional community (Figure 5A). However, some OTUs were rare in the regional community according to the DNA data set but showed intermediate abundances within the RNA data set and vice versa. Approximately 25 OTUs were disproportionately underrepresented by RNA tags (Figure 5B, gray area), suggesting low activity or

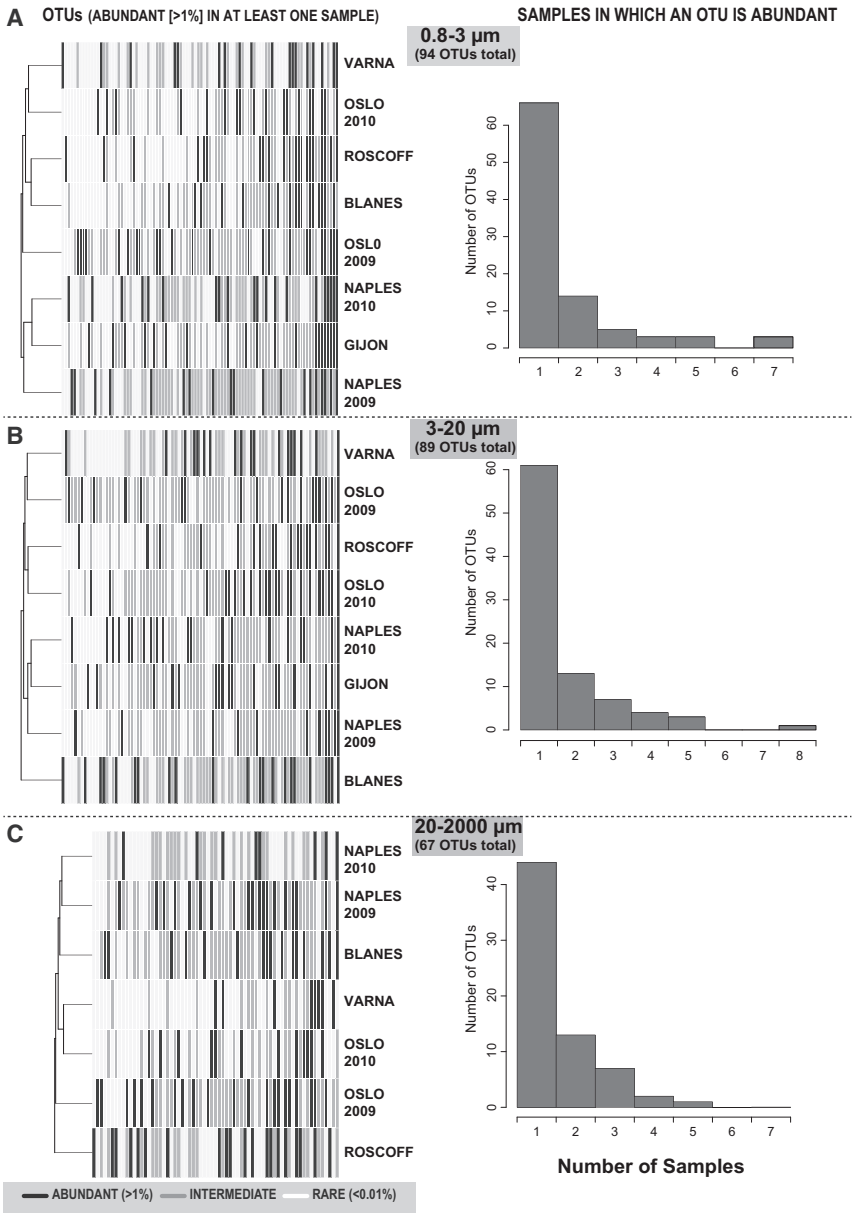


Figure 3. Locally Abundant OTUs Tended to Be Abundant in a Single Sample

(A–C) Abundance across samples/sites of OTUs that were locally abundant in at least one sample, separated by size fractions. Heatmaps (left) indicate whether OTUs (vertical lines) were abundant (>1%; black) or rare (<0.01%; white) or had intermediate (gray) abundances in specific samples/locations. Histograms on the right indicate the number of samples in which OTUs were abundant in each data set.

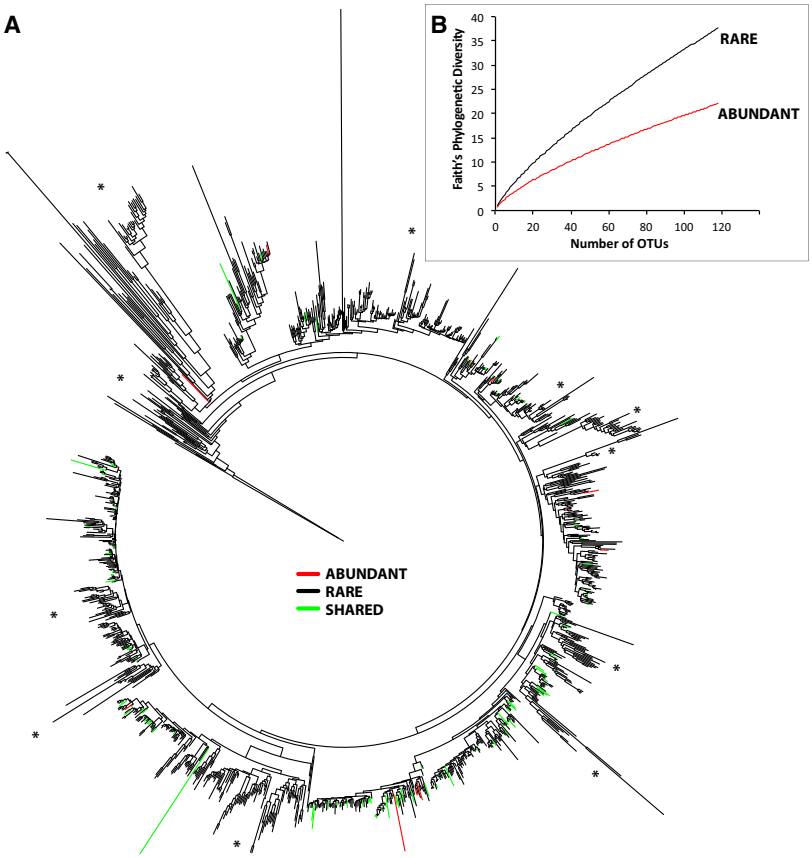
large numbers of rDNA copies in the genome. On the contrary, ~20 OTUs were disproportionately overrepresented by RNA tags, which may result from high ribosomal activity (Figure 5B, yellow area).

Discussion

Marine microbial eukaryotes constitute arguably the most poorly characterized biodiversity component in the biosphere [20]. Here, we provide new insights into the structural and phylogenetic organization of their communities by using the first ultra-deep sequencing data set of 18S rRNA tags extracted from surface picoplankton, nanoplankton, and microplankton/mesoplankton collected from six marine coastal locations across Europe. Illumina sequencing of >150 million V9 rRNA amplicons followed by highly stringent sequence quality filtering allowed us to approach richness saturation (OTUs 95%) in both the entire regional community and in most local communities, thus allowing the exploration of the rare protist biosphere. The highest richness was observed in the proximity of the smallest cell sizes (in the nanoplankton), thus resembling patterns observed in animals by early ecologists [34, 35]. We estimate a recovery of ~64% of the total number of OTUs in the entire region; therefore, more samples are required to cover the total diversity of European coastal waters. As observed in aquatic prokaryotes [9], most of the recovered OTUs (>63.0%) belonged to the rare biosphere. Because we used RNA as template, we can attest that these OTUs represent living, ribosomically active cells.

Despite the strong spatiotemporal variability characterizing marine coastal waters and the different β-diversity among sites, the proportion of locally rare and abundant taxa was remarkably constant across all sampled communities. This pattern suggests community self-organization arising from local species interactions, with the observed regular proportions representing stable community configurations [36]. Given this striking consistency observed in our data, we hypothesize that in other marine planktonic communities, >70% of protist OTUs are rare as well. However, note that rarity was analyzed according to one pre-existing definition; future studies should explore multiple definitions in order to determine which one is the most meaningful [10].

Both the abundant and rare regional communities demonstrated contrasting patterns regarding their general structure. The abundant regional community was predominantly structured by organismal size fraction, whereas the rare regional community was structured mostly according to geographic site. On the one hand, size fraction structuring reflects the fact that—excluding protists with complex cell cycles, ontogenic processes, or cell shapes and colony forms markedly distinct from a sphere—most taxa have rather constant cell sizes. On the other hand, site-associated clustering indicates that the differences among communities from different sites are larger than the differences among size fractions within the same site. Such groupings of the rare picoplankton, nanoplankton, and microplankton/mesoplankton were generated by only a few OTUs that were present in only one site and



C

	# OTUs	$MPD_{observed}$	MPD_{null}	MPD_z	*p*	$MNTD_{observed}$	$MNTD_{null}$	$MNTD_z$	*p*
Regionally Abundant OTUs	118	0.94	1.25	-4.34	< 0.001	0.22	0.39	-4.62	< 0.001
Regionally Rare OTUs	1332	1.24	1.24	-0.1	0.386	0.19	0.19	-0.76	0.160

Figure 4. Phylogenetic Patterns in the Abundant and Rare Regional Communities

(A) Maximum-likelihood phylogenetic tree based on reference Sanger 18S sequences (n = 1,343) representing regionally abundant and rare OTUs (each branch of the tree derives from a reference sequence that represents an OTU). Red indicates sequences representing only regionally abundant OTUs, black points to sequences exclusively representing regionally rare OTUs, and green indicates sequences representing both regionally abundant and rare OTUs. Asterisk indicates some groups that were formed entirely by rare OTUs.

(B) Rarefaction analysis of Faith's phylogenetic diversity considering abundant and rare OTUs in the regional community.

(C) MPD and MNTD estimates based on the phylogeny shown in (A). $MPD_{observed}$ denotes observed MPD values. MPD_{null} denotes MPD values obtained from a null model. MPD_z denotes standardized effect size of MPD = ($MPD_{observed}$ − $MPD_{null\ model}$)/sd($MPD_{null\ model}$); p denotes p value. The same parameters are shown for the MNTD. Negative MPD_z and $MNTD_z$ values with $p < 0.05$ indicate phylogenetic overclustering.

Note that in (B) and (C), sequences that represented both abundant and rare OTUs ("shared") were considered to be within both the abundant and rare data sets.

were shared between the large and smaller size fractions. These OTUs could represent low-abundance protists with life cycles that involve different cell sizes, different lifestyles (host associated or not), and issues related to a nonoptimal size fractionation during filtering. Site-associated clustering can be promoted by historical contingencies occurring in different communities, such as local random extinctions or stochastic immigration events [37], which are expected to have a larger impact on rare subcommunities, making them generally more distinct among each other than their abundant counterparts.

Even though abundant and rare regional communities presented a markedly different general structure, we found a moderate but significant correlation in their β diversity that points to similar biogeography for some rare and abundant taxa. This suggests that similar structuring processes can affect both abundant and rare subcommunities and that the rare protist biosphere is not a random collection of taxa. Comparable results have been reported for marine and lacustrine prokaryotes [17, 19, 38].

Underlying the β diversity patterns at the regional level, locally abundant (>1%) OTUs within the picoplankton, nanoplankton, and microplankton/mesoplankton showed marked variations in relative abundance among samples. Most locally abundant OTUs were abundant in only one sample, having intermediate or low (<0.01%) abundances in the other samples. In addition to reflecting strong fluctuations in protistan abundance across heterogeneous coastal locations or seasonality in the same site [23, 39], this pattern points to a general decoupling between local and regional abundances because most OTUs that are abundant in only one location will not be regionally abundant.

Finally, the rare protistan biosphere presented a distinctive phylogenetic composition, with a significant proportion of rare OTUs phylogenetically unrelated to abundant ones. In particular, several clades contained exclusively rare OTUs that were relatively distantly related in phylogenetic terms to the nearest abundant taxon. Although we cannot ignore the fact that some taxa from these exclusively rare clades may be abundant in other locations or seasons, the overall pattern suggests that permanent or semipermanent rarity (achieved, e.g., through a low cell-division rate) may be an evolutionary trait of some marine protist groups. Avoidance of competition, predation, and parasitism are potential advantages of a low-abundance life [8], which could evolve through negative frequency-dependent selection [40]. On the contrary, rare OTUs in the regional community that were phylogenetically closely related to abundant ones could represent intragenomic variation or erroneous variants of abundant OTUs generated during PCR or sequencing [41, 42], although we minimized this bias by working with a relatively loose definition of OTUs at 95% similarity threshold [27]. The structuring of the abundant regional community seems to have been influenced by environmental selection of evolutionary-related taxa presenting favorable traits because abundant taxa were phylogenetically more closely related than expected by chance [32, 33]. Our comparison of rRNA- versus rDNA-derived OTUs indicated that both types of markers are broadly

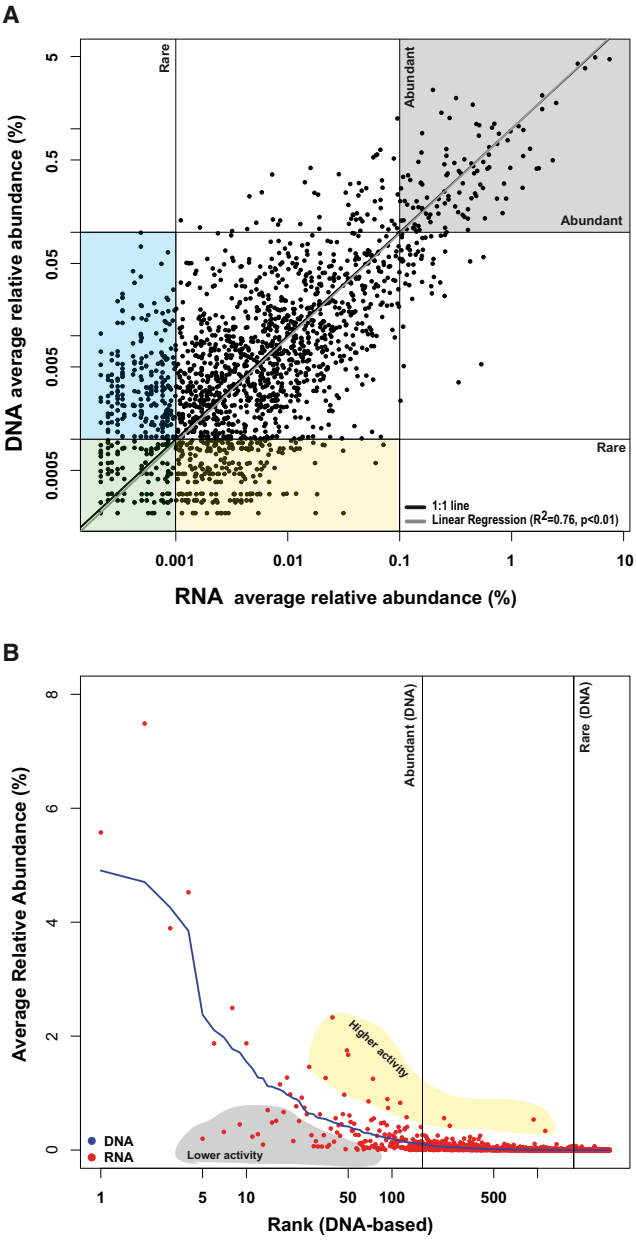


Figure 5. Abundance of OTUs in the Regional Community According to rDNA and rRNA

(A) Average relative abundance of individual OTUs (dots) according to rRNA and rDNA. The abundance thresholds for abundant (>0.1%) and rare (<0.001%) are indicated with vertical and horizontal lines. The top right gray corner indicates OTUs that were abundant in both RNA and DNA, whereas the bottom left green corner indicates OTUs that were rare in both RNA and DNA. The yellow section indicates OTUs that were rare according to DNA and not rare according to RNA, and the light blue section indicates the opposite. The best-fitting linear regression, which was virtually identical to the 1:1 line, is indicated.

(B) OTU rank-abundance curve based on rDNA (blue line) and the corresponding abundance for each OTU according to rRNA (red dots). Abundant (>0.1%) and rare (<0.001%) thresholds are indicated with vertical lines. OTUs disproportionately overrepresented or underrepresented in RNA, in comparison to DNA, are indicated in the yellow and gray areas, respectively.

See also Table S4.

positively correlated, supporting the hypothesis that low metabolic activity or dormancy is not common among planktonic microbial eukaryotes [14, 43]. Thus, metabolically active taxa likely prevail in the protistan rare planktonic biosphere. In addition, rRNA/rDNA comparisons suggested that disproportionately high activity is unusual in planktonic protists. Altogether, this contrasts markedly with planktonic bacteria, in which dormancy appears to be more prevalent [9, 14] and in which, for some taxa, activity can increase as abundance decreases [15].

Conclusion

Overall, our results indicate that marine planktonic protist communities are composed of predominantly active abundant and rare subcommunities with contrasting structuring patterns and phylogenetic characteristics, which nevertheless display striking consistency in their local relative proportions, even in dissimilar coastal waters. Further analyses of protist community structuring in contrasting oceanic biomes will provide a wider test of the patterns we found in European coastal waters, contributing altogether to a better understanding of the community organization mechanisms in microbial eukaryotes and their links to local and global ecosystem functioning.

Experimental Procedures

Sampling and Illumina or 454 Sequencing

Surface (<5 m depth) seawater samples were collected from six European coastal offshore sites: Blanes (Mediterranean), Gijon (Bay of Biscay), Naples (Mediterranean), Oslo (North Sea/Skagerrak), Roscoff (English Channel), and Varna (Black Sea) (Figure 1A; Table S3). Picoplankton and nanoplankton samples were collected by using Niskin bottles. A total of 15–40 l of water was prefiltered through a 20 μm sieve and then sequentially filtered through polycarbonate membranes of 3 μm and 0.8 μm. Microplankton/mesoplankton samples were collected using a 20-μm-porosity plankton net, concentrating the samples by filtering water for 20 min. Then they were prefiltered through a 2,000 μm sieve and afterward filtered through a 20 μm polycarbonate membrane. Total RNA and DNA were extracted simultaneously from the three membranes. For Illumina GAIIx sequencing, hypervariable V9 18S tags were PCR amplified from cDNA obtained from RNA template, whereas V4 18S tags were PCR amplified from both DNA and RNA (cDNA) templates for 454-Titanium sequencing.

Sequence Analysis for Illumina Reads

A total of 23 samples was selected for downstream analyses (Table S1). For the forward reads (hereafter referred to as "reads"), about 15 Gb of raw sequence data (100 bp reads) was produced (Table S1). Reads (minimum 90 bp) were quality checked by using a sliding 10 bp window, and each window had to have a Phred quality average > 34 to pass the control. The number of clean reads after quality control is shown in Table S1. Quality-checked reads were analyzed in QIIME [44] version 1.4. Reads were clustered into OTUs by using UCLUST version 1.2.22 [45] with a 95% similarity threshold. Chimeras were detected by using ChimeraSlayer [46], with a reference database derived from PR2 [47]. Taxonomy assignment was done by comparing, using BLAST [48], the most abundant (representative) sequence of each OTU against different reference databases, and unwanted OTUs (e.g., metazoa and prokaryotes) were removed. The final curated Illumina RNA data set included 5,696,049 reads.

Sequence Analysis for 454 Reads

We analyzed 15 samples for which both DNA and RNA V4 18S tags were sequenced (Table S4); these samples were also present in the Illumina data set (Table S1). All 454 reads between 200 and 500 bp were run through QIIME version 1.4. Reads were checked for quality by using a sliding window of 50 bp (Phred average > 25 in each window) and truncated to the last good window. Sequences were denoised by using DeNoiser version 0.851 [49], as implemented in QIIME version 1.4, and then clustered into OTUs by using UCLUST version 1.2.22 with a 99% similarity threshold. Chimera detection and taxonomy assignment were done by using the

same approaches as with the Illumina reads. In the final V4 curated data set, RNA included 233,085 reads and DNA included 221,898 reads (454,983 reads total).

Final OTU Tables

Single singletons and OTUs present in a single sample were removed from both Illumina V9 and 454 V4 OTU tables. For both data sets, we randomly subsampled OTU tables to the number of reads present in the sample with the lowest amount of reads. This value was 78,000 reads per sample for Illumina and 3,000 reads per sample for 454.

Ad Hoc Definitions of Rare and Abundant OTUs

OTUs were classified as abundant or rare in relation to their local and regional relative abundances. Locally abundant OTUs were defined as those with relative abundances >1%, and locally rare OTUs were defined as those with abundances <0.01%, following studies in prokaryotes [9, 17] and protists [25]. Regional relative abundances for specific OTUs were calculated as the average of local relative abundances for such OTUs across all samples, including zero values. The thresholds for defining abundant and rare at the regional level were arbitrarily defined as the local thresholds divided by a factor of ten. OTUs abundant in the regional community had a mean relative abundance of >0.1%, whereas regionally rare OTUs had a mean relative abundance of <0.001%.

Diversity Analyses

Most analyses were run in R [50] environment by using the packages Vegan [51] and Picante [52]. Rarefactions and species (OTUs) accumulation curves were calculated in Vegan. OTU networks were constructed in QIIME based on the subsampled OTU table and graphically edited in Cytoscape [53] using the layout "edge-weighted spring embedded" with eweights.

Mapping of Illumina Reads to Reference Sanger Sequences and Phylogeny Construction

Representative reads of regionally abundant or rare OTUs were mapped separately to a custom V9 18S rDNA Sanger reference database based on the PR2 [47] by using BLASTn. We used an e value $<1 \times 10^{-6}$ with a percentage of identity >90% to assign all abundant (n = 154) OTUs and 95% of the rare (n = 5,329) OTUs to reference taxa. The chosen parameters allowed for different OTUs to be mapped to the same Sanger reference taxa, and, for this reason, the final phylogeny had fewer taxa than the sum of abundant and rare OTUs. For phylogeny construction, we extracted the full-length 18S sequence corresponding to all reference V9. Sequences were aligned by using Mothur against the aligned SILVA 108 database (eukaryotes only). A maximum-likelihood reference tree (8,311 sequences) was calculated by using RAxML HPC-MPI version 7.2.8 [54] under the model GTR+CAT/G+I and checked against other phylogenies of marine protists [55] for consistency. The tree was pruned using the R package analyses of phylogenetics and evolution (APE) [56] to keep only those reference taxa that were hit by abundant or rare OTUs. We used the final pruned tree, including 1,343 Sanger sequences, to calculate the MPD and MNTD [32] with Picante. Phylogenetic diversity [31] was computed by using Picante.

See more details on experimental procedures in Supplemental Experimental Procedures.

Accession Numbers

The accession numbers for the Illumina and 454 sequences reported in this paper are 4549916.3–4549968.3 and are publicly available at MG-RAST (http://metagenomics.anl.gov/).

Supplemental Information

Supplemental Information includes Supplemental Experimental Procedures, five figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2014.02.050.

Acknowledgments

We thank the Biodiversity of Marine euKaryotes (BioMarKs, http://www.biomarks.eu) consortium, which was funded by the European Union ERA-Net program BiodivERsA (2008-6530). Extra financial support was provided by the Marie Curie Intra-European Fellowship (PIEF-GA-2009-235365) and Juan de la Cierva (JCI-2010-06594) programmes to R.L. and FLAME (CGL2010-16304, MICINN, Spain) to R.M. The Barcelona Supercomputing Center (BSC) provided access to the MareNostrum supercomputer (grants BCV-2011-2-0003/3-0005 and 2012-1-0006/2-0002 to R.L. and R.M.). We thank N. Le Bescot for assistance with Figure 1 and the two reviewers who helped to improve this work.

Received: December 29, 2013
Revised: January 27, 2014
Accepted: February 20, 2014
Published: April 3, 2014

References

1. Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008). The microbial engines that drive Earth's biogeochemical cycles. Science *320*, 1034–1039.
2. DeLong, E.F. (2009). The microbial ocean from genomes to biomes. Nature *459*, 200–206.
3. Massana, R. (2011). Eukaryotic picoplankton in surface oceans. Annu. Rev. Microbiol. *65*, 91–110.
4. Caron, D., Countway, P., Jones, A., Kim, D., and Schnetzer, A. (2012). Marine protistan diversity. Ann Rev Mar Sci. *4*, 467–493.
5. Logares, R. (2006). Does the global microbiota consist of a few cosmopolitan species? Ecología Austral *16*, 85–90.
6. Martiny, J.B., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., et al. (2006). Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. *4*, 102–112.
7. Arrigo, K.R. (2005). Marine microorganisms and global nutrient cycles. Nature *437*, 349–355.
8. Pedrós-Alió, C. (2006). Marine microbial diversity: can it be determined? Trends Microbiol. *14*, 257–263.
9. Pedrós-Alió, C. (2012). The rare bacterial biosphere. Annu. Rev. Mar. Sci. *4*, 449–466.
10. Magurran, A.E., and McGill, B.J. (2011). Biological Diversity: Frontiers in Measurements and Assessment (Oxford: Oxford University Press).
11. Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc. Natl. Acad. Sci. USA *103*, 12115–12120.
12. Logares, R., Haverkamp, T.H., Kumar, S., Lanzén, A., Nederbragt, A.J., Quince, C., and Kauserud, H. (2012). Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. J. Microbiol. Methods *91*, 106–113.
13. Caron, D., and Countway, P. (2009). Hypotheses on the role of the protistan rare biosphere in a changing world. Aquat. Microb. Ecol. *57*, 227–238.
14. Jones, S.E., and Lennon, J.T. (2010). Dormancy contributes to the maintenance of microbial diversity. Proc. Natl. Acad. Sci. USA *107*, 5881–5886.
15. 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. USA *108*, 12776–12781.
16. Sjöstedt, J., Koch-Schmidt, P., Pontarp, M., Canbäck, B., Tunlid, A., Lundberg, P., Hagström, A., and Riemann, L. (2012). Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. Appl. Environ. Microbiol. *78*, 1361–1369.
17. Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. Proc. Natl. Acad. Sci. USA *106*, 22427–22432.
18. 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.
19. Logares, R., Lindström, E.S., Langenheder, S., Logue, J.B., Paterson, H., Laybourn-Parry, J., Rengefors, K., Tranvik, L., and Bertilsson, S. (2013). Biogeography of bacterial communities exposed to progressive long-term environmental change. ISME J. *7*, 937–948.
20. Caron, D.A., Worden, A.Z., Countway, P.D., Demir, E., and Heidelberg, K.B. (2009). Protists are microbes too: a perspective. ISME J. *3*, 4–12.
21. Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M.J., Chistoserdov, A., Orsi, W., and Edgcomb, V.P. (2009). Massively

parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. BMC Biol. *7*, 72.
22. Cheung, M.K., Au, C.H., Chu, K.H., Kwan, H.S., and Wong, C.K. (2010). Composition and genetic diversity of picoeukaryotes in subtropical coastal waters as revealed by 454 pyrosequencing. ISME J. *4*, 1053–1059.
23. Nolte, V., Pandey, R.V., Jost, S., Medinger, R., Ottenwälder, B., Boenigk, J., and Schlötterer, C. (2010). Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. Mol. Ecol. *19*, 2908–2915.
24. Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D., 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* (*Suppl 1*), 21–31.
25. Mangot, J.F., Domaizon, I., Taib, N., Marouni, N., Duffaud, E., Bronner, G., and Debroas, D. (2013). Short-term dynamics of diversity patterns: evidence of continual reassembly within lacustrine small eukaryotes. Environ. Microbiol. *15*, 1745–1758.
26. Sieburth, J.M., Smetacek, V., and Lenz, J. (1978). Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fractions. Limnol. Oceanogr. *23*, 1256–1263.
27. Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. *12*, 118–123.
28. Magurran, A.E. (2004). Measuring Biological Diversity (Oxford: Blackwell Publishing).
29. Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Automat. Contr. *19*, 716–723.
30. Preston, F.W. (1948). The commonness, and rarity, of species. Ecology *29*, 254–283.
31. Faith, D. (1992). Conservation evaluation and phylogenetic diversity. Biol. Conserv. *61*, 1–10.
32. Webb, C.O., Ackerly, D.D., McPeek, M.A., and Donoghue, M.J. (2002). Phylogenies and community ecology. Annu. Rev. Ecol. Syst. *33*, 475–505.
33. Cavender-Bares, J., Kozak, K.H., Fine, P.V., and Kembel, S.W. (2009). The merging of community ecology and phylogenetic biology. Ecol. Lett. *12*, 693–715.
34. Hutchinson, G.E., and MacArthur, R.H. (1959). A theoretical ecological model of size distributions among species of animals. Am. Nat. *93*, 117–125.
35. Rosenzweig, M.L. (1995). Species Diversity in Space and Time (Cambridge: Cambridge University Press).
36. Zimmermann, C.R., Fukami, T., and Drake, J.A. (2003). An experimentally-derived map of community assembly space. Y. Bar-Yam and A. Minai, eds. Unifying Themes in Complex Systems II: Proceedings of the Second International Conference on Complex Systems, 427–436.
37. Ricklefs, R.E. (1987). Community diversity: relative roles of local and regional processes. Science *235*, 167–171.
38. Gobet, A., Böer, S.I., Huse, S.M., van Beusekom, J.E., Quince, C., Sogin, M.L., Boetius, A., and Ramette, A. (2012). Diversity and dynamics of rare and of resident bacterial populations in coastal sands. ISME J. *6*, 542–553.
39. Kim, D.Y., Countway, P.D., Jones, A.C., Schnetzer, A., Yamashita, W., Tung, C., and Caron, D.A. (2014). Monthly to interannual variability of microbial eukaryote assemblages at four depths in the eastern North Pacific. ISME J. *8*, 515–530.
40. Hibbing, M.E., Fuqua, C., Parsek, M.R., and Peterson, S.B. (2010). Bacterial competition: surviving and thriving in the microbial jungle. Nat. Rev. Microbiol. *8*, 15–25.
41. Huse, S.M., Welch, D.M., Morrison, H.G., and Sogin, M.L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environ. Microbiol. *12*, 1889–1898.
42. Quince, C., Lanzen, A., Davenport, R.J., and Turnbaugh, P.J. (2011). Removing noise from pyrosequenced amplicons. BMC Bioinformatics *12*, 38.
43. Massana, R., and Logares, R. (2013). Eukaryotic versus prokaryotic marine picoplankton ecology. Environ. Microbiol. *15*, 1254–1261.
44. Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.
45. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.
46. Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., et al.; Human Microbiome Consortium (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. *21*, 494–504.
47. Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., et al. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. *41* (Database issue), D597–D604.
48. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.
49. Reeder, J., and Knight, R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat. Methods *7*, 668–669.
50. R Development Core Team (2008). R: A language and environment for statistical computing (Vienna: R Foundation for Statistical Computing).
51. Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Wagner, H. (2008). Vegan: community ecology package. R package version 1.15-0 (Finland: University of Oulu).
52. Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., and Webb, C.O. (2010). Picante: R tools for integrating phylogenies and ecology. Bioinformatics *26*, 1463–1464.
53. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.
54. Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.
55. Pernice, M.C., Logares, R., Guillou, L., and Massana, R. (2013). General patterns of diversity in major marine microeukaryote lineages. PLoS ONE *8*, e57170.
56. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. Bioinformatics *20*, 289–290.