environmental microbiology reports

Environmental Microbiology Reports (2011) **3**(6), 661–666 doi:10.1111/j.1758-2229.2011.00250.x

Repeated sampling reveals differential variability in measures of species richness and community composition in planktonic protists

John R. Dolan[1]* and Thorsten Stoeck[2]

[1]*Microbial Ecology and Biogeochemistry, Laboratoire d'Océanographie de Villefranche-sur-Mer, Université Paris 6 CNRS UMR 7093, Observatoire Océanologique de Villefranche-sur-Mer, Station Zoologique, B.P. 28,F-06230 Villefranche-sur-Mer, France.*

[2]*University of Kaiserslautern, Ecology Department, Faculty of Biology, Erwin-Schrödinger Str. 14, D-67663 Kaiserslautern, Germany.*

Summary

Diversity metrics and descriptors of protistan community structure were calculated from 12 samples of 10 l each collected from the Bay of Villefranche in the NW Mediterranean Sea. Variability of the sampling was on scales of minutes and meters. The individual samples were compared with each other and compared with a pooled data set from the total volume of 120 l, considered as the 'true' community. We focused on a single group of planktonic protists, tintinnids, a coherent functional and phylogenetic group in which morpho-species identifications by light microscopy are unambiguous. Tintinnid abundance in the samples ranged from 217 to 321 cells of 16–21 species with the number of rare species in a sample (< 1% of abundance) positively related to species richness of the sample. Rarefaction estimates of total species richness in the 12 samples ranged from 21 ± 3.5 to 37 ± 3.6 compared with the 34 species of the pooled data set. The measures of similarity reflected the differences between samples in both the numbers and identities of the least abundant or rare species. The species abundance distribution using pooled data was best fit by a log-series or geometric distribution; eight species accounted for about 90% of total cells and most species, the remaining 22 out of 34, were 'rare' (concentration < 1% of total cells). Among the samples, 5 were best fit by a geometric model, 1 by a log-series distribution, 2 by a log-normal or log-series model, and 4 could not be clearly assigned a particular distribution. Our results suggest that single sample estimates of species richness are relatively robust compared with measures of taxonomic similarity and species abundance distribution. When measuring differences among populations sample variability should be considered.

Received 23 November 2011, accepted 29 January, 2011. For correspondence. *E-mail dolan@obs-vlfr.fr; Tel. (+33) 4 93 76 38 22; Fax (+33) 4 93 76 38 34.

Introduction

In recent years, many studies have reported on differences in species richness and taxonomic composition across space and time in planktonic protist communities using a bewildering variety of techniques (Heywood *et al*., 2010). In particular pyrosequencing is now being touted (Dawson and Hagen, 2009; Creer, 2010) and used to describe a dramatic decline in protist diversity with depth in the North Pacific (Brown *et al*., 2009), diversity in a temperate estuary as five times that of Antarctic water (Amaral-Zettler *et al*., 2009), and as invariant, in terms of species richness but not composition, among different seasons in a lake (Nolte *et al*., 2010) or spatially distant anoxic water bodies (Stoeck *et al*., 2009). The conclusions reached by these and other studies are difficult to evaluate not only as each technique has its own limits and biases (compounded by study-specific variants), but because many, if not most, recent studies of microbial diversity do not report variability among replicate samples (Prosser, 2010).

We were interested in investigating the variability of different measures of protist diversity while avoiding contentious problems of technological biases and limits. We employed direct microscopic examination of material from repeated sampling and restricted analysis to a single, coherent, group of organisms in which species identifications are relatively unambiguous. We focused on the well-known, circumscribed, coherent group of planktonic protist, tintinnid ciliates. Tintinnids are all herbivorous microzooplankton, and thus belong to the same functional group and are phylogentically united within a single monophyletic ciliate order, *Tintinnidia* (Lynn, 2008). All have a shell (lorica) into which the ciliate cell can withdraw. Characteristics of the lorica, visible using light microscopy, are used to distinguish species (or

morpho-species) and are also well-correlated with a variety of ecological characteristics of the species (Dolan, 2010).

We analysed samples from 12 l Niskin bottle casts at 25 m depth; the samples were taken over a period of 45 min during which time the boat drifted a maximum of 200 m from the first sampling location. Variability of the sampling was then on scales of minutes and meters, similar to that of replicate sampling taken using multiple bottles on a single large Niskin bottle rosette, or repeating a Niskin bottle cast. The variability between samples was 'fine-scale' and fine-scale patches are known to characterize the distributions of planktonic protists, specifically ciliates, diatoms and nanoflagellates (e.g. Montagnes *et al.*, 1999). The goal was to estimate variability in metrics of protist community structure: estimates of species richness, taxonomic distinctness, and patterns of species abundance. We sought to empirically assess the robustness of these metrics through repeated sampling of planktonic protists, known to exhibit patchy distributions. To our knowledge, no previous study has addressed this issue. Microscopic examination was used to enumerate all tintinnid cells in the samples and provided 12 distinct data sets for comparison among each other and a pooled data set taken to yield the best estimates of the 'true' or complete community.

Results and discussion

The complete data set is available as an Excel file as supplementary material (see Material S1). Plotting species accumulation curves for the 12 samples revealed similar patterns for the first 2 l examined yielding 11–14 species. Subsequently, the curves diverged to give a total of 16–21 species for the next 10 l of the sample. While plateaus were approached, none of the accumulation curves showed a clear plateau, including that of the pooled sample (Fig. 1), suggesting that even 120 l was insufficient to encounter all of the tintinnid species. These apparently undetected species would be part of the 'rare' tintinnid community, those present in very low concentrations of a few individuals per m^3.

Summary data for the 12 samples and pooled data are given in Table 1. The total number of cells varied from 217 to 321 and species richness ranged from 16 to 21. Tintinnid abundance in a sample was unrelated either to species richness or to the numbers of 'rare' species in a sample (species representing < 1% of cell abundance). However, the species richness of a sample was positively related to the number of rare species found in the sample (Fig. 2).

The taxonomic composition of assemblages was similar in that 80–90% of individuals in each sample (see Material S1 for species names and abundances) were composed of only eight abundant species. However, the qualitative Jaccard index of similarity, based on presence/absence data, showed wide variability in average similarity scores both between samples as well as to the pooled data set (Table 1). The Jaccard similarity between any given sample and the pooled data set was inversely related to the samples' average similarity to all the other individual samples (Fig. 3). This means, the higher the Jaccard similarity between an individual subsample is to

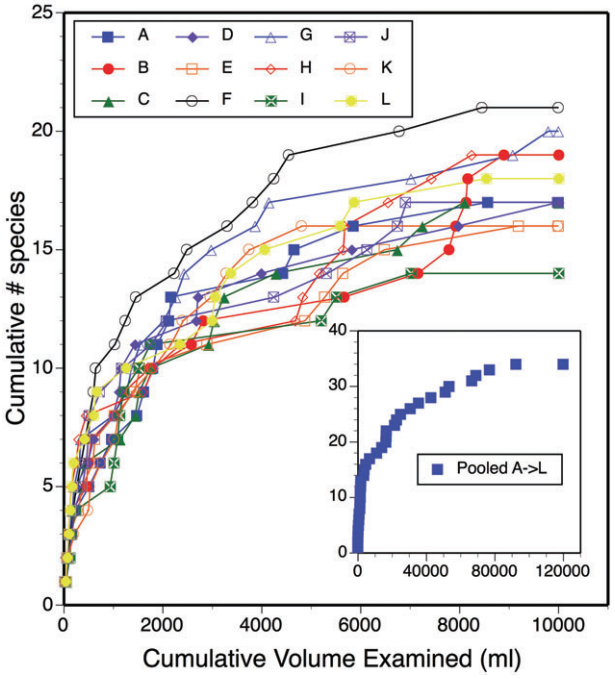

Fig. 1. Species accumulation curves of 12 samples and a pooled data set representing samples A through L. Samples are from the standard sampling station 'Point B' (43°41′10″N, 7°19′00″E, approximately 100 m total depth), at the entrance to the Bay of Villefranche-sur-Mer in the NW Mediterranean Sea. On 29 June 2010, A 12 l Niskin bottle (integrates ≈ 1 m of depth) was used to sample at 25 m depth. The 12 bottle casts took about 45 min. During the sampling, the boat drifted off Point B by a maximum of 200 m. Thus, sample variability of the 12 samples was on scales of minutes and metres, similar to that of replicate sampling taken using multiple bottles on a single Niskin bottle rosette, or repeating Niskin bottle casts. In the laboratory each 10 l sample was concentrated to 20 ml by slowly and gentle screening through a 20 μm Nitex screen fixed to the bottom of a 10 cm dia. PVC tube. Concentrated samples were fixed with alkaline Lugol's (2% final concentration). Samples were arbitrarily labelled A through L, not reflecting the actual sampling order. Processing of the 12 samples took approximately 1 h. Fixed samples were stored refrigerated until microscopic analysis which was completed within 3 months of sampling. For each sample, the entire concentrate was examined at 200× using an inverted microscope in 2 ml aliquots settled in sedimentation chambers. Tintinnid species assignations were made based on lorica morphology using standard taxonomic monographs (see Dolan *et al.*, 2009). Empty lorica were not enumerated. For each sample, in addition to noting abundances of each species found, a running sum of cells encountered was maintained and the first occurrence of each species noted to furnish data for construction of species accumulation curves.

Table 1. Summary data and similarity indices of the 12 samples and the pooled set of data.

Sample	Number of cells	Species richness	No. of rare species	Avg spp Shared	Avg Jaccard	Avg Morisita–Horn	Jaccard w/pooled	Morisita–Horn w/pooled
A	217	17	6	12.9	0.600	0.904	0.500	0.942
B	288	19	7	13.1	0.567	0.934	0.599	0.977
C	257	17	6	12.3	0.561	0.928	0.500	0.972
D	251	16	5	13.0	0.650	0.905	0.471	0.949
E	273	16	6	13.0	0.632	0.93	0.471	0.968
F	233	21	9	13.5	0.550	0.928	0.618	0.971
G	297	20	7	14.4	0.632	0.883	0.558	0.925
H	331	19	9	13.2	0.575	0.929	0.599	0.971
I	258	14	4	12.2	0.630	0.900	0.412	0.945
J	244	17	6	12.8	0.598	0.942	0.500	0.98
K	282	15	3	12.9	0.663	0.943	0.441	0.982
L	321	18	7	13.7	0.641	0.928	0.529	0.97
Pooled	3252	34	22	17.4	0.511	0.962	–	–

Rare species were those found in concentrations of < 1% of the cells in the particular sample, or for pooled, the entire data set. We employed the simple Jaccard index of taxonomic overlap and the preferred Morista-Horn index which accounts for proportional representation of taxa (Magurran, 2004). All samples were compared with all other samples and average values calculated for numbers of shared species (Avg spp Shared), average values of the qualitative Jaccard index (Avg Jaccard) and the quantitative Morisita–Horn index (Avg Morisita–Hom) are given. Each sample was also compared with the pooled data set using the Jaccard (Jaccard w/pooled) and Morisita–Horn (Morisita–Horn w/pooled) indices.

Fig. 2. Scatterplot of the number of species found in an individual sample and the number of rare species (< 1% of total cells in the sample). Data from the pooled sample 'P' are plotted only for illustration. The regression equation is $y = 0.762x - 7.02$; $r^2 = 0.795$; $n = 12$; $P < 0.01$.

Fig. 3. Scatterplot of the indices of similarity of each sample to the pooled sample and the average similarity compared with the other 11 samples (see Table 1). The qualitative presence/absence Jaccard index indicated that similarity of a individual sample to the pooled sample is inversely related to average similarity to other samples ($y = -0.4x + 0.81$; $r^2 = 0.474$; $n = 12$; $P < 0.05$). Thus, a sample distinct from other samples, in terms simply of the species found, was more likely to resemble the pooled sample. In contrast, the quantitative Morisita–Horn index, which accounts for relative abundances, showed that similarity to the other samples and the pooled data set was positively related ($y = 1.0x - 0.07.$; $r^2 = 0.983$; $n = 12$; $P < 0.01$).

Table 2. Results of rarefaction analysis of the 12 samples and pooled data.

Sample	ACE Mean	ACE SD	Chao 1 Mean	Chao 1 95% CI Low	Chao 1 95% CI High	Chao 1 SD	Cole rarefaction	Cole SD
A	22.21	4.8	20.56	18.07	35.79	3.46	22.46	1.99
B	29.02	5.45	26.88	23.72	44.02	4.06	26.64	1.82
C	33.61	7.68	30.68	26.98	49.04	4.49	28.9	1.63
D	36.72	6.74	33.58	29.8	52.47	4.62	30.32	1.48
E	35.68	4.03	34.85	31.5	52.18	4.17	31.3	1.34
F	35.77	3.57	36.02	32.66	53.5	4.2	32.04	1.2
G	35.39	2.8	36.18	33.11	52.5	3.89	32.61	1.05
H	35.56	1.83	36.63	33.76	51.67	3.61	33.06	0.9
I	35.31	1.37	35.79	33.99	45.86	2.36	33.41	0.73
J	35.17	1.29	35.15	34.09	41.54	1.47	33.67	0.56
K	34.85	1.16	34.86	34.15	39.11	0.98	33.86	0.37
L	34.59	1.07	34.52	34.1	37.03	0.6	33.96	0.19
Pooled	34	0	34	34	34	0	34	0

Note that the different methods yielded estimates close to the richness of the pooled data set except for samples A and B. Species Richness Estimates were calculated using EstimateS (Colwell, R.K. 2009: EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.2. User's Guide and application, http://purl.oclc.org/estimates).

the pooled data set, the lower is the similarity of this subsample to the other samples. Thus, a sample that is distinct from other samples, in terms of the presence/absence of species, was more likely to resemble the pooled sample. The differences can be explained in terms of 'rare species'. The number of rare species in a sample was positively related to its similarity to the pooled sample ($y = 0.03x + 0.3$; $n = 12$; $P = 0.01$) and negatively related to average similarity to other samples ($y = -0.014x + 0.7$; $n = 12$, $P = 0.05$), in terms of the Jaccard index. The Morisita–Horn values, a quantitative measure of similarity that accounts for the relative abundances of different species, were on average high (Table 1), reflecting the fact that the samples had in common not only the eight abundant species but also their relative abundances. For a given sample, similarity to the pooled data was positively related to similarity to the other samples (Fig. 3). Overall, the measures of similarity reflected the differences between samples in the numbers and identities of the least abundant or rare species.

Rarefaction analysis using different methods yielded estimates of species richness across the samples of near parity (Table 2). The average number of species with the different methods was quite close to that of 34 in the pooled data, and ranged narrowly with Cole rarefaction yielding 31 species, Chao1 33 species, and ACE 33.7 species. Interestingly, two samples A and B appeared to be relative 'outliers' in species richness (Table 2) but were not distinct from other samples in taxonomic similarity (Table 1). For a given sample, the accuracy of the species richness estimates (deviation from the 34 of the pooled data) was unrelated to its similarity either to the pooled data or to other samples. Thus, robustness of the species richness estimate for a sample was independent of its taxonomic representivity.

Patterns of species abundance are thought to reflect the factors regulating, or determining, relative abundances of different species. A geometric species abundance distribution represents monopolization of resources by species arriving sequentially (Whittaker, 1972); log-series corresponds to results of random dispersal from a larger community of ecologically similar forms (Hubbell, 2001); log-normal species abundance distribution is produced when species are limited by multiple factors acting multiplicatively and is the most common in species-rich assemblages (Magurran, 2004).

The species abundance distribution of the pooled data set, examined through comparisons with model-derived distributions, was most closely fit by either a log-series or geometric distribution. Among the sample populations the patterns of species abundance varied considerably but the geometric was the most common (Table 3). Out of the 12 individual populations, 5 appeared most closely fit by a geometric distribution, 1 by a log-series distribution, 2 by a log-normal or log-series model. For four samples, a pattern could not be assigned, the fits of log-normal, geometric or log-series were not clearly distinguishable from one another. The apparent variability found in species abundance distributions was due to the fact that individual samples represented an incomplete species census. Notably, under-sampled populations are often best fit by a power law (e.g. geometric) distribution (Ulrich *et al.*, 2010).

Overall, the variability between samples produced minor differences in species richness estimates as most closely approached that of the pooled data set. However, the samples showed various distributions of species abundance, which suggests distinct community structures among sample populations. Furthermore, the individual samples, while all dominated by the same eight abundant

Table 3. Results of the analysis of species abundance distributions.

Sample	Log-normal fit	Geometric fit	Log-series fit
A	**1.657**	2.361	**1.251**
B	2.055	**–0.198**	1.568
C	1.761	**–1.189**	1.330
D	**1.687**	**1.163**	**1.297**
E	**1.914**	**1.676**	**1.687**
F	1.529	1.915	**0.540**
G	**1.758**	2.747	**1.012**
H	2.200	**0.764**	1.797
I	1.743	**–1.230**	1.594
J	**1.779**	**1.296**	**1.386**
K	**1.348**	**1.371**	**1.144**
L	1.987	**0.515**	1.585
Pooled	3.225	**2.631**	**2.159**

For each sample, and the pooled data set, the log-rank abundance curve of each of the assemblage was compared with model-derived log-normal, geometric and log-series curves using the Akaike Information Criterion (AIC) test. The values in bold denote the closest fit as indicated by the lowest AIC value; differences below 1 (multiple values in bold) indicate indistinguishable fits (Burnham and Anderson, 2002). Note that the pooled data set showed a geometric or log-series distribution while among the samples, 5 were best fit by a geometric model, 1 by a log-series distribution, 2 by a log-normal or log-series model, and 4 could not be clearly assigned a particular distribution.
To examine patterns of species abundance distribution we constructed log-rank abundance curves for each sample by calculating relative abundance for each species and ranking species from highest to lowest and plotting ln(relative abundance) versus rank. Then, we constructed hypothetical log-rank abundance curves that could fit the data by using parameters of the particular assemblage. We constructed curves for three different popular models of community organization: geometric series, log-series and log-normal, as in Dolan and colleagues (2007; 2009) and Raybaud and colleagues (2009). The observed rank abundance distributions were compared with the hypothetical models using a Bayesian approach: an Akaike Goodness of fit calculation (Burnham and Anderson, 2002). Using this approach, an AIC was determined as the natural logarithm of the mean (sum divided by S) of squared deviations between observed and predicted ln(relative abundance) for all ranked S species plus an additional term to correct for the number of estimated parameters, k (1 for geometric series and 2 each for log-series and log-normal distributions): $(S + k)/(S - k - 2)$. The lower the calculated AIC value, the better the fit. A difference of 1 in AIC corresponds roughly to a 1.5 evidence ratio; we considered differences of less than 1.0 in AIC to represent indistinguishable fits following Burnham and Anderson (2002, pp. 142–143).

species, showed considerable taxonomic distinction. These differences, whether relative to the pooled data set or to other individual samples, were closely related to the number and identity of 'rare species' present in a sample. It is worth recalling that incomplete sampling of a population in which most species are 'rare' will yield various numbers of rare species. We found that despite enumerating all the cells in a sample of relatively large volume (10 l), a single sample under-censused the tintinnid population. Individual samples contained only 40–60% of the species in the pooled data set (Fig. 1). Thus, individual sample populations appeared to be taxonomically distinct largely because most species (22 out of 34) were rare, occurring at concentrations of less than 1 per litre in our samples. It is worthwhile recalling that tintinnids are not the only planktonic protists which occur in relatively low concentrations of 10^0–10^1 l^{-1}, such abundances also characterize radiolarians, foraminifera and many species-rich groups of dinoflagellates among others. This may in part explain reports in recent observations in molecular protistan surveys in which the majority of taxa in protistan communities occur in very low abundances (Amaral-Zettler *et al*., 2009; Stoeck *et al*., 2009; Nolte *et al*., 2010).

Our results argue for the employment of repeated and intensive sampling. Different species and assemblages of species likely exhibit distinct patterns of variability reflecting spatial and temporal patchiness. For example, among planktonic ciliates in a tropical lagoon, patches of different sizes characterize different species of oligotrichs and tintinnids (Bulit *et al*., 2003) and consequently ciliate diversity can show fine-scale patchiness (Bulit *et al*., 2009). In contrast, in Mediterranean tintinnid communities examined by pooling several large samples over the water column and sampled over time scales of days, both of which yield raw counts of 10^3 organisms, communities appear relatively coherent (e.g. Dolan *et al*., 2009). We can speculate then that had we analysed larger samples, variability between samples would have declined. While sampling strategies can take into account the time and space scales of the targeted organisms, patchy distributions are an 'inconvenient truth' that requires recognition.

Our observations have clear implications with regard to investigations focusing on 'rare species' which are beginning to receive close attention (e.g. Caron and Countway, 2009; Dawson and Hagen, 2009). For example, a recent study of seasonal changes in the planktonic protist community in a lake (Nolte *et al*., 2010) reported the existence of a stable set of abundant taxa coupled with a highly variable set of seasonally changing 'rare species'. The rarefaction analysis of single monthly samples (Nolte *et al*., 2010: Fig. 1) suggested consistent under-sampling, similar to our study (Fig. 1). As patchiness can yield apparent under-sampling, it is difficult to distinguish actual differences between populations from sample variability (e.g. Youssef *et al*., 2010). Similarly, conclusions of differences among spatially separated populations, such as comparing a surface water to a deep sea protist community (Brown *et al*., 2009) although intuitively reasonable, still require some consideration of sample variability (Prosser, 2010). Similarly, the question of microbial endemicism or cosmopolitanism cannot be addressed without knowledge of the representivity of the samples used.

Acknowledgements

This study was conducted in the framework of the Aquaparadox project financed by the Agence National de Recherche programme 'Biodiversité and the Pôle Mer PACA. T.S.

acknowledges the support of the Deutsche Forschungsgemeinschaft (DFG) through Grant STO414/3-1 and funds from the University of Kaiserslautern. Mark E. Ritchie (Syracuse University) provided spreadsheets to model and compare species abundance distributions and Todd W. Arnold (University of Minnesota) provided advice on determining model fits using the AIC criterion. The aid of Jean-Yves Carval in obtaining the samples as well as the constructive and encouraging comments of the reviewers and editor are gratefully acknowledged.

References

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. doi:10.1371/journal.pone.0006372.

Brown, M.V., Philip, G.K., Bunge, J.A., Smith, M.C., Bisset, A., Lauro, F.M., *et al*. (2009) Microbial community structure in the North Pacific Ocean. *ISME J* **3:** 1374–1386.

Bulit, C., Diaz-Avalos, C., Signoret, M., and Montagnes, D.J.S. (2003) Spatial structure of planktonic ciliate patches in a tropical coastal lagoon: an application of geostatistical methods. *Aquat Microb Ecol* **30:** 185–196.

Bulit, C., Diaz-Avalos, C., and Montagnes, D.J.S. (2009) Scaling patterns of plankton diversity: a study of ciliates in a tropical coastal lagoon. *Hydrobiologia* **624:** 29–44.

Burnham, K.P., and Anderson, D.R. (2002) *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. New York, USA: Springer.

Caron, D.A., and Countway, P.D. (2009) Hypotheses on the role of the protistan rare biosphere in a changing world. *Aquat Microb Ecol* **57:** 227–238.

Creer, S. (2010) Second-generation sequencing derived insights into temporal biodiversity dynamics of freshwater protists. *Mol Ecol* **19:** 2829–2831.

Dawson, S.C., and Hagen, K.D. (2009) Mapping the protistan 'rare biosphere'. *J Biol* **8:** 105.

Dolan, J.R. (2010) Morphology and ecology in tintinnid ciliates of the marine plankton: correlates of lorica dimensions. *Acta Protozoologica* **49:** 235–344.

Dolan, J.R., Ritchie, M.R., and Ras, J. (2007) The 'neutral' community structure of planktonic herbivores, tintinnid ciliates of the microzooplankton, across the SE Tropical Pacific Ocean. *Biogeosciences* **4:** 297–310.

Dolan, J.R., Ritchie, M.E., Tunin-Ley, A., and Pizay, M.-D. (2009) Dynamics of core and occasional species in the marine plankton: tintinnid ciliates in the north-west Mediterranean Sea. *J Biogeogr* **36:** 887–895.

Heywood, J.L., Sieracki, M.E., Bellows, W., Poulton, N.J., and Stepanauskas, R. (2010) Capturing diversity of marine heterotrophic protists: one cell at a time. *ISME J* (in press).

Hubbell, S.R. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ, USA: Princeton University Press.

Lynn, D.H. (2008) *The Ciliated Protozoa: Characterization, Classification, and Guide to the Literature*. Berlin, Germany: Springer.

Magurran, A.E. (2004) *Measuring Biological Diversity*, 2nd edn. Oxford, UK: Blackwell Publishing.

Montagnes, D.J.S., Poulton, A.J., and Shammon, T.M. (1999) Mesoscale, finescale and microscale distribution of micro- and nanoplankton in the Irish Sea with emphasis on ciliates and their prey. *Mar Biol* **134:** 167–179.

Nolte, V., Pandey, R.V., Jost, S., Medinger, R., Ottenwalder, B., Boenigk, J., and Schlotterer, C. (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Mol Ecol* **19:** 2908–2915.

Prosser, J.L. (2010) Replicate or lie. *Environ Microbiol* **12:** 1806–1810.

Raybaud, V., Tunin-Ley, A., Ritchie, M. E., and Dolan, J.R. (2009) Similar patterns of community organization characterize distinct groups of different trophic levels in the plankton of the NW Mediterranean Sea. *Biogeosciences* **6:** 431–438.

Stoeck, T., Behnke, A., Christen, R., Amaral-Zettler, L., Rodriguez-Mora, M.J., Chistoserdov, A., *et al*. (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biol* **7:** 72.

Ulrich, W., Ollik, M., and Ugland, K.I. (2010) A meta-analysis of species-abundance distributions. *Oikos* **119:** 1149–1155.

Whittaker, R.H. (1972) Evolution and measurement of species diversity. *Taxon* **21:** 213–251.

Youssef, N.H., Couger, M.B., and Elshahed, M.S. (2010) Fine-Scale Bacterial Beta Diversity within a Complex Ecosystem (Zodletone Spring, OK, USA): The Role of the Rare Biosphere. *PLoS ONE* **5:** e12414.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Material S1. Count data for the 12 samples.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.