ORIGINAL ARTICLE

Patterns and processes in microbial biogeography: do molecules and morphologies give the same answers?

Luciana F Santoferrara¹, Jean-David Grattepanche², Laura A Katz^{2,3} and George B McManus¹

¹Department of Marine Sciences, University of Connecticut, Groton, CT, USA; ²Department of Biological Sciences, Smith College, Northampton, MA, USA and ³Program in Organismic and Evolutionary Biology, University of Massachusetts, Amherst, MA, USA

Our knowledge on microbial biogeography depends on the way we define and study diversity. In contrast to most microbes, some protist lineages have conspicuous structures that allow comparisons of diversity concepts and measures—those based on molecules and those based on morphology. We analyzed a group of shell-bearing planktonic ciliates, the tintinnids, in a coast-to-ocean gradient using high-throughput sequencing and microscopy. First, we compared molecular operational taxonomic units (OTUs) and morphospecies in terms of assemblage composition, distribution and relationships with the environment. OTUs revealed potentially novel and rare taxa, while morphospecies showed clearer correlations with environmental factors, and both approaches coincided in supporting a coastal versus oceanic pattern. Second, we explored which processes influence assembly across the environmental gradient examined. Assemblage fluctuations were associated with significant distance–decay and changes in morphospecies size and prey proxies, thus suggesting niche partitioning as a key structuring mechanism. Our conclusion is that molecules and morphologies generally agreed, but they provided complementary data, the first revealing hidden diversity, and the latter making better connections between distribution patterns and ecological processes. This highlights the importance of linking genotypes and phenotypes (using multidisciplinary analyses and/or reliable databases of barcoded species), to understand the diversity, biogeography and ecological roles of microbes.

The ISME Journal (2016) 10, 1779–1790; doi:10.1038/ismej.2015.224; published online 5 February 2016

Introduction

A central debate in microbial biogeography has contrasted species distributions as either ubiquitous or restricted in space and time, but the coexistence of both patterns is now well supported (for example, van der Gast, 2015). In addition to observations of cosmopolitan distributions for some species, multiple examples of endemism have been found for terrestrial and aquatic archaea, bacteria and eukaryotic microbes, including protists (Foissner, 2008; Fontaneto, 2011). Even marine planktonic microbes, which are thought to have high dispersal potential due to large population sizes and the apparent lack of barriers in the ocean, include both globally-distributed and geographically-isolated taxa (Pomier *et al.*, 2007; Casteleyn *et al.*, 2010). However, a major factor influencing our understanding of biogeography is the way we define and study taxa

(Hanson *et al.*, 2012), thus raising the question whether our current perceptions are real or semantic.

In protists, two main characteristics have been used to track taxon distributions: morphologies and molecules. As opposed to bacteria and archaea, many protist taxa have structural traits that allow species identification. For instance, some marine protists build shells with morphological features that have allowed taxonomic classification by standard microscopy (for example, foraminiferans, radiolarians and tintinnid ciliates). In these groups, morphologically-defined species (morphospecies) have historically been found to display distinct distribution patterns in both the horizontal and vertical dimensions (for example, Boltovskoy, 1999). Nevertheless, biogeographical patterns based on morphospecies may be obscured by cryptic and polymorphic species, given that conserved, convergent or plastic morphologies can mask major molecular, biological, physiological and ecological differences (Dolan, 2015). The increasing combination of morphological and DNA sequence analyses in single specimens has revealed, for instance, cryptic species with different distributions and likely dissimilar ecology in marine plankton (Weiner *et al.*, 2012;

Correspondence: LF Santoferrara, Department of Marine Sciences, University of Connecticut, 1080 Shennecossett Road, Groton, CT 06340, USA.

E-mail: luciana.santoferrara@uconn.edu

Received 19 June 2015; revised 28 October 2015; accepted 3 November 2015; published online 5 February 2016

Ishitani *et al.*, 2014; Santoferrara *et al.*, 2015). Furthermore, the characterization of natural communities by high-throughput sequencing (HTS) is now offering the potential to reveal new distribution patterns based on the detection of rare species and the discovery of novel, atypical taxa that may not be recognized by microscopy, even in some groups with a long tradition of morphological description (Lecroq *et al.*, 2011). At the same time, these conspicuous groups are useful to test known issues of environmental HTS, as parallel morphospecies identification in the microscope helps to differentiate meaningful operational taxonomic units (OTUs based on sequence similarity) from potential HTS errors (for example, Bachy *et al.*, 2013; Santoferrara *et al.*, 2014).

We use tintinnid ciliates as a model to compare morphologies and molecules. The lorica is the main character used to delimit tintinnid morphospecies, which agree partially with OTUs defined by ribosomal DNA sequences (Santoferrara *et al.*, 2013). Tintinnids are major consumers of phytoplankton, and their feeding ecology relates to morphology given that prey size is determined by the diameter of the lorica aperture (Dolan *et al.*, 2013a, b). Biogeographical patterns based on worldwide morphological data are known for about half of the 75 extant genera, including both restricted (coastal versus oceanic and warm-temperate versus polar) and cosmopolitan examples (Dolan and Pierce, 2013). Indeed, assemblages of tintinnid morphospecies differ markedly in the bathymetric, latitudinal and vertical profiles (for example, Alder, 1999; Modigh *et al.*, 2003; Santoferrara and Alder, 2012), as well as in the seasonal cycle of temperate coasts (for example, Bojanić *et al.*, 2012). Structuring of morphospecies assemblages has been explained by environmental selection in a coastal site (Sitran *et al.*, 2009) or random dispersal in open waters (Dolan *et al.*, 2007, 2009, 2013b), but the processes that affect assembly in the transitions between environments are unknown. Likewise, it is not known whether patterns and processes based on morphologies and molecules agree, as few distribution studies have used DNA sequences to target tintinnids and aloricate sister lineages (Doherty *et al.*, 2010; Tamura *et al.*, 2011; Bachy *et al.*, 2014; Grattepanche *et al.*, 2015).

We studied the fluctuation of the tintinnid assemblage in a coast-to-ocean gradient of the NW Atlantic Ocean (sites from ca. 30 to 1000 m deep) using microscopy and HTS. First, we questioned whether morphospecies and OTUs agree in terms of assemblage composition, distribution patterns and correlation with environmental variables. Given the complex relationship between morphologies and molecules, our hypothesis was that the two approaches would provide different answers. Second, we explored which ecological processes may explain assembly in the transition between coast and ocean, under the hypothesis that the environmental gradient across shelf waters selects assemblage members.

Materials and methods

Sampling

We sampled 23 stations regularly separated by 7.4 km in a transect from 13 to 176 km off the coast of Rhode Island (USA) during 6–9 July 2012 on board the R/V Cape Hatteras (Figure 1a; Grattepanche *et al.*, 2015). Vertical profiles of temperature, salinity, oxygen concentration and chlorophyll fluorescence were measured with a Seabird CTD profiler mounted on a rosette. Seawater was collected with Niskin Bottles at four depths: surface, pycnocline, chlorophyll maximum depth (CMD) and deepest (ca. 5 m above the bottom, except for stations 21–23 sampled at 100–150 m). From the CMD, 250 ml of seawater was concentrated on glass microfiber GF/F filters, which were stored in the dark at –80 °C until high-performance liquid chromatography (HPLC) analysis of phytoplankton accessory pigment composition and concentration. Three pigment-based ratios were estimated as proxies for the relative abundances of micro-, nano- and pico-phytoplankton (Supplementary Table S1).

From each of the four depths sampled, two series of samples were collected for ciliates. For microscopy, 250 ml of seawater was preserved with acid Lugol's solution (2% f.c.). For molecular analysis, 1 l of seawater was screened through an 80-µm mesh to remove metazoa, and then sequentially filtered through 10 µm and 2 µm polycarbonate filters, which were stored in 0.5 ml of DNA buffer at 4 °C. Additionally, surface samples collected with a 20-µm-mesh net were preserved with non-acid Lugol's solution (2% f.c.) for single-cell sequencing. To estimate the biomass of potential predators at each station, a 150-µm-mesh net was towed vertically from the deepest waters reached by the rosette to the surface. Samples were fixed with formalin (10% f.c.), collected on glass fiber filters in duplicate, dried at 60 °C for 24 h and weighed.

Microscopy and single-cell sequencing

Tintinnids were studied with an inverted microscope. For each of 91 acid Lugol's samples, 150 ml was settled and entirely surveyed. Loricae containing their cell were counted, measured and identified following original descriptions and the work of Alder (1999). Morphospecies found in non-acid Lugol's samples were isolated and sequenced using the method by Santoferrara *et al.* (2013). The small subunit ribosomal DNA (SSU rDNA) sequences from 10 morphospecies, 5 of them newly sequenced, were deposited in NCBI GenBank (accession codes KT792924 to KT792933; Supplementary Results).

HTS and sequence analysis

HTS focused on 28 samples corresponding to the four depths sampled at seven stations regularly distributed across the continental shelf (Figure 1a; Supplementary Table S2). DNA was extracted from one-half of each

Figure 1 The environmental gradient analyzed. (a) Stations analyzed by microscopy only (circle) or microscopy and HTS (square; modified from Grattepanche *et al.*, 2015). (b) Principal component analysis (PCA) showing a coast-to-ocean gradient (represented by clearer-to-darker colors) at four depth layers. For each environmental variable, eigenvectors and their highest values on the corresponding PCs are shown. (c–e) HPLC-estimated chlorophyll *a* concentration and phytoplankton size fraction ratios at the CMD, and mesozooplankton biomass in the water column, respectively. Linear fits and significant Pearson's coefficients (*R*, *P* < 0.05) against the distance to the coast are shown.

filter and only samples from the 10-μm filters were used in this study given the prevalence of known tintinnid species above this size. Library preparation, HTS, sequence quality filtering and OTU identification were based on published methods (Santoferrara *et al.*, 2014; see details in the Supplementary Methods). In brief, a SSU rDNA region was amplified with primers specific for tintinnids, other choreotrichs and oligotrichs (Tamura *et al.*, 2011). Then, five independent PCR products per sample were pooled and sequenced (454 Life Sciences, Roche, Branford, CT, USA). Pyrosequences were deposited in NCBI Sequence Read Archive (accession code SRP064473).

QIIME v. 1.8.0 (Caporaso *et al.*, 2010) was used for quality filtering and OTU clustering. OTU clustering was done *de novo* using the UCLUST algorithm (Edgar, 2010) and a 100% similarity cutoff. Chimeras were identified *de novo* with UCHIME as implemented in USEARCH (Edgar *et al.*, 2011) and removed. OTUs derived from only five or fewer reads, as well as OTUs that occurred in only one sample, were excluded as potential errors. Non-tintinnid OTUs were removed (Supplementary Methods).

A self-curated reference database of morphospecies sequences was used to identify putative tintinnid OTUs in two approaches. First, the tintinnid OTUs were combined with the reference sequences and used in a phylogenetic approach. Sequences of aloricate choreotrichs, oligotrichs and stichotrichs were added as outgroups. An alignment was done with MUSCLE (Edgar, 2004), and refined with

Clustal Omega (Sievers *et al.*, 2011). Positions containing more than 70% gapped characters were removed. A Maximum Likelihood tree was built with RAxML (Stamatakis, 2006) using 1000 bootstrap replicates and the GTR model of evolution with a gamma model of rate heterogeneity and a proportion of invariable sites, as previously identified with jModelTest under the Akaike Information Criterion (Darriba *et al.*, 2012). In the second approach, the tintinnid OTUs were contrasted against the reference sequences using BLASTN (Altschul *et al.*, 1990). Unidentified environmental sequences in GenBank were excluded from the analyses, except for the BLASTN comparison of OTU groups that did not include any known species.

Sample comparison and multivariate analyses

To compare HTS data among samples, the number of sequences per sample was standardized to the minimum obtained (4783 sequences; station 4, surface) using a random sub-sampling without replacement in QIIME (Caporaso *et al.*, 2010). Standardization was not needed among microscopy samples, because identical volumes were always examined. However, inherent technical differences prevent accurate standardization of sampling effort between HTS and microscopy (Supplementary Figure S1; see below).

Multivariate analyses of HTS data were done with QIIME (Caporaso *et al.*, 2010) or with R using the

vegan package (Oksanen *et al.*, 2014); the latter was used also for microscopy and environmental data. Ordination of all samples based on normalized environmental data (temperature, salinity, oxygen concentration and chlorophyll fluorescence) was done with a principal components analysis (PCA) based on Euclidean distance. Correlation between the distance to the coast and HPLC-estimated chlorophyll *a* concentration and phytoplankton size fraction ratios, as well as mesozooplankton biomass, was tested with the Pearson's coefficient. Ordination of samples based on tintinnids was done by non-metric multidimensional scaling using Bray-Curtis dissimilarity matrices of either morphospecies or OTUs. Additional ordinations were done by principal coordinates analysis based on either weighted or unweighted Unifrac distance matrices of OTUs. Clustering of samples based on tintinnids was done with each of these matrices using the unweighted pair-group method with arithmetic averages and 1000 replicates for jackknife analyses. The main morphospecies and OTUs driving Bray-Curtis dissimilarities among sample groups were identified.

Overall correlation of Bray-Curtis β -diversity and matrices of pairwise environmental distance (normalized Euclidean distance), horizontal distance (between stations) or vertical distance (between depths sampled) were tested using simple and partial Mantel tests with 10 000 permutations. Significance levels were corrected for multiple comparisons. The best subset of environmental variables producing the maximum correlation with β -diversity was estimated (BIOENV test).

Results and Discussion

Environmental conditions

The environmental variables we measured showed a gradient from coastal to oceanic waters (Figure 1). The PCA showed continuous changes both horizontally and in the vertical profile (PC1 and 2, respectively; Figure 1b). These changes were linked mainly to a decrease in chlorophyll fluorescence and oxygen concentration from inshore to offshore, a decrease in temperature from surface to the deepest waters, and an increase in salinity in the two dimensions (Figure 1b). Distance to the coast was significantly correlated with a decrease in chlorophyll *a* concentration and changes in phytoplankton size fraction proxies (the microphytoplankton ratio decreased as the nanophytoplankton ratio increased) from inshore to offshore in the CMD (Figures 1c and d) as well as with a decrease in mesozooplankton biomass in the water column (Figure 1e). These trends suggest a productivity decrease toward open waters and a switch in phytoplankton size fractions in relation to the vertical structure of the water column (Supplementary Figure S2A), in agreement with known dominance of larger cells in coastal mixed waters and smaller cells offshore (for example, Li, 2002).

Comparison of tintinnid morphospecies and OTUs

By contrasting morphospecies and OTUs in a coast-to-ocean gradient from surface to bottom waters, we provide a framework to explore some of the strengths and weaknesses of HTS in environmental surveys and expand the spatial scale of tintinnid assemblage comparisons by modern and classical methods (Bachy *et al.*, 2013; Santoferrara *et al.*, 2014). An inherent limitation of such comparisons is the difficulty in standardizing sampling effort between methods due to the different steps involved in each procedure (Supplementary Figure S1). For microscopy, a precisely measured volume of sample is settled and examined (generally based on cell density and work effort). For HTS, an unknown portion of the sample volume ends contributing to the final data set: only a fraction of extracted DNA is used for non-quantitative PCR, then only a fraction of amplified DNA is used for HTS, and then only a fraction of sequences is kept after bioinformatic quality filtering. Although HTS is (ideally) designed for representativeness and sensitivity, the volume actually examined is not exactly comparable to the one studied by microscopy, even starting with identical sample sizes. This prevents the strict comparison of α -diversity values between both techniques (Magurran, 2003), which is also limited by other factors (for example, different susceptibility to cell destruction during preservation or filtration; Santoferrara *et al.*, 2014). Still, we were able to extract valuable information by evaluating the overall performance of each method and comparing a) assemblage composition, b) distribution patterns and c) correlations with environmental variables in terms of morphospecies and OTUs.

a) Assemblage composition. We found a taxonomic composition that was less diversified in terms of morphospecies than OTUs (Figure 2). Among the 28 samples analyzed by both approaches, morphospecies and OTUs shared four taxonomic groups (*Steenstrupiella*/*Amphorides*, *Salpingella*/*Amphorellopsis*, *Eutintinnus* and other tintinnids), but there were four additional OTU groups. Of these OTU groups, three may correspond to microscopically-observed species with unknown sequences or to novel taxa (S1, S2 and Eu); the fourth is known (*Tintinnidium*) but was detected in very low proportion (0.04% reads). Furthermore, of the 30 morphospecies detected by microscopy, 17 have known SSU rDNA sequence and all of them were found also by HTS. There were no cases of morphospecies with known sequence detected by microscopy but not by HTS. In contrast, of the 509 putative tintinnid OTUs detected, 23 are identical to sequences of known morphospecies, the 17 above and 6 additional OTUs (<0.2% reads) that were not found by microscopy (Figure 2, Supplementary Tables S4 and S5). Although differences in species detection may be related to sample volumes

(Supplementary Figure S1), our HTS approach showed high sensitivity for tintinnid studies.

Relative abundances agree only partially for morphospecies and OTUs (Figure 2). Nine

morphospecies and nine OTUs were dominant ($\geq 3\%$ cells or reads, respectively), but only four of them were common to both approaches (black arrows in Figure 2). An effect of either species

Figure 2 Tintinnid assemblage composition by microscopy and HTS. *Left*, sketch of a RAXML tree based on sequences of known morphospecies and pyrosequenced OTUs (Supplementary Figure S4). Underlined species were newly sequenced (Supplementary Results, Supplementary Figure S3, Supplementary Table S3). Circles indicate nodes with bootstrap support > 50%, although the short sequence analyzed mostly prevents statistically-relevant inferences. Morphospecies of known sequence detected by both HTS and microscopy are highlighted in yellow, based on both identity in the RAXML tree and 100% similarity in BLASTN (¹99.7% similarity, equivalent to one base substitution). BLASTN results are shown for three groups of OTUs that did not include any sequence in GenBank (despite their similarity to known genera, these groups may belong to other taxa according to the RAXML tree) and for the only dominant OTU with unknown morphology but 100% similar to an environmental GenBank sequence. *Right*, bubble plots representing relative abundance of cells or pyrosequences (values $\geq 3\%$ are indicated). Arrows indicate taxa with contribution $\geq 3\%$ in both approaches (black), only by microscopy (white), or only by HTS (gray). Microscopy results are shown for ²28 samples also studied by HTS and ³91 total samples. Not detected (-), unknown sequence or morphology (?).

dimensions compared with the pore size used for sample filtering or primer mismatches is not evident, but there could be other PCR bias (for example, selective amplification of closely-related species), taxon-specific susceptibility to DNA extraction or variability in the copy number of the rDNA operon

(Medinger *et al.*, 2010; Heywood *et al.*, 2011). This supports known inconsistencies in relative abundances estimated by HTS and microscopy for ciliates and other protists (Bachy *et al.*, 2013; Egge *et al.*, 2013; Santoferrara *et al.*, 2014; Stoeck *et al.*, 2014).

The total number of OTUs we found (Figure 2) represents almost half of the number of described tintinnid species (ca. 1200). Although many of the unidentified OTUs may be real (for example, cryptic or rare species not detected in the microscope), others may be artifactual variants of the dominant OTUs, which cause richness inflation in the HTS data. Each dominant OTU (generally correlated to the distribution of their corresponding morphospecies, if known) was associated with multiple, closely-related OTUs (that is, same cluster in RAxML tree and similarity >99%) that usually occurred in the same samples but with many fewer reads (Figures 3a–e). These variants may result from the known intra-individual and intra-specific variability in tintinnid SSU rDNA (Gong *et al.*, 2013; Santoferrara *et al.*, 2013) or from HTS errors (Reeder and Knight, 2009), in part retained in the data due to our OTU clustering at 100% similarity. A lower cutoff is not adequate for our amplicon because that can cause sequences of separate known species to be clustered into the same OTU (Grattepanche *et al.*, 2014; Santoferrara *et al.*, 2014).

Our 100% OTU clustering strategy, however, did allow for the detection of potentially novel diversity within known genera. For example, closely-related *Salpingella* OTUs that had high abundances in distant samples could represent different species (Figure 3f). These OTUs could not be linked with the *Salpingella* spp. observed in the microscope, because their sequence is unknown. Furthermore, some morphospecies with small loricae (diameter 4–14 µm) were difficult to differentiate, did not match any described species (Supplementary Figure S3, Supplementary Table S4) and were probably underestimated by both microscopy (Figure 2) and HTS (as we used DNA samples from 10-µm filters). Also in the Mediterranean, where tintinnids have been characterized by morphology and molecules, some *Salpingella* OTUs are only detected by environmental sequencing (Bachy *et al.*, 2013). Thus, *Salpingella* may include closely-related species that cannot be discriminated by light microscopy. Cryptic species with non-overlapping distributions possibly linked to niche separation have been detected in tintinnids

Figure 3 Occurrence of dominant OTUs and their corresponding closely-related OTUs. Some identified OTUs co-occurred with multiple, much less abundant closely-related OTUs (a–d). OTU-4245 from the group S2 co-occurred with all but 12 of its closely-related OTUs (e). In the case of *Salpingella* OTUs (f), four dominant ($\geq 3\%$ reads) and two moderately-abundant (1–2% reads) OTUs were closely-related but had different distributions (OTU-4040 was found in most samples, OTU-6989 peaked in the sample 7d and OTUs 5650, 3982, 2810 and 4106 were detected only in stations 12–23); there were also multiple, scarce, close OTUs. The distribution of dominant OTUs was also compared with their corresponding morphospecies, if known (r = Spearman's coefficient; $P < 0.01^{**}$).

(Santoferrara *et al.*, 2015) and may be more common than expected.

b) Distribution patterns. Morphospecies and OTUs captured similar patterns of spatial distribution: inshore and offshore sample groups were detected by all the analyses done (Figure 4, Supplementary Figure S6). A deep sample group was detected using morphospecies and non-quantitative OTU data (Figures 4a and c; Supplementary Figure S6A and C), but it was not shown by quantitative OTU data (Figures 4b and d; Supplementary Figure S6B and D). The inconsistent detection of the deep group (mostly the deepest samples from stations 14 to 23) may be related to the lower number of samples examined by HTS compared with microscopy, although an influence of sample reduction was not expected given that morphological findings were very similar for the 28 samples studied by both methods and the total 91 samples (Figure 2). A more likely explanation for differences is the impact of

biased HTS quantification on β -diversity (Lozupone and Knight, 2008). Using phylogenetic information, unweighted (non-quantitative) and weighted (quantitative) Unifrac ordinations were significantly correlated (Procrustes analysis, $m^2 = 0.75$, $P = 0.001$), but the latter was less consistent even in separating inshore and offshore samples (Figures 4c and d). Bray-Curtis (non-phylogenetic, quantitative) analyses showed an intermediate performance, with inshore and offshore separation in the ordination analysis (Figure 4b), but less support according to cluster analysis (Supplementary Figure S6B). Despite the differences, two-thirds of the samples were always assigned to the same group regardless of the approach used (Supplementary Figure S7).

Tintinnid assemblages clearly differed among sample groups, despite some differences in the identities of the main morphospecies and OTUs driving the groupings (Figure 5). These results were driven by the most abundant taxa and reflected the previously mentioned differences in relative

Figure 4 Ordination of samples based on morphospecies and OTUs. Results were compared with cluster analyses (Supplementary Figure S6) to establish sample groups: inshore (green), offshore (light blue) and deep (dark blue; not detected in **b** and **d**). Outliers and missing samples correspond to very low occurrence (< 5 cells) or non-detection of tintinnids, respectively. (**a**, **b**) Non-metric multidimensional scaling based on Bray-Curtis similarity matrices of morphospecies and OTUs, respectively. (**c**, **d**) Principal coordinates analysis based on unweighted and weighted Unifrac distance matrices of OTUs, respectively. Samples examined by both microscopy and HTS or only by microscopy are indicated by black and gray circles, respectively. Samples are labeled with station number and depth (s = surface, p = pycnocline, c = chlorophyll maximum, d = deepest). In (**a**), a part of the ordination was magnified (incomplete line; see all labels in Supplementary Figure S5). In (**c**, **d**), the proportion of variance explained by each axis (Pc) is shown.

abundances between microscopy and HTS (Figure 2). However, because the most abundant taxa were also the most frequent within each group, they were still found as the main contributors to the grouping using presence/absence data (Figure 5). This supports the idea that rare OTUs (that is, with low abundance and occurrence, such as many of our potential artifacts) have a weak effect on β -diversity detection, unless they are studied separately from the abundant OTUs (Zinger *et al.*, 2012; Logares *et al.*, 2014; Lynch and Neufeld, 2015). Also, given that the detection of rare taxa is the most affected by sample size, presence/absence data suggest that our β -diversity findings were not strongly impacted by

the different sample volume examined by microscopy and HTS (Supplementary Figure S1).

Our cross-shelf data supported patterns of tintinnid species ubiquitous in both coastal and open waters, as well as species typical from only one type of environment (Alder, 1999; Dolan and Pierce, 2013). Some morphospecies and OTUs occurred both inshore and offshore, and even in the deepest waters we examined (for example, *Eutintinnus perminutus* and the *Salpingella* OTUs 4040 and 6989; Supplementary Table S6). This, and a pattern of distinct inshore and offshore distributions, agrees with the findings for the most abundant choreotrich and oligotrich ciliates in the same

Figure 5 Assemblage composition in inshore, offshore and deep waters. (a, b) Proportion of cells and pyrosequences by morphospecies and OTUs, respectively, in each sample group (see values in Supplementary Table S6). Numbers on the bars indicate the proportion of group dissimilarity attributable to the relative abundance (or, in brackets, presence/absence) of each morphospecies or OTU. Only values > 5% for OTU presence/absence or > 10% for the rest are shown.

samples (Grattepanche *et al.*, 2015). The tintinnid morphospecies or OTUs we found inshore were nearly or completely undetectable offshore, and vice versa (Supplementary Table S6), with a boundary at approximately 35 km from shore, near the 50-m isobath (Station 4, Figure 1a). Limited distributions on the scale we analyzed (10 s of km) do not mean that these species are spatially restricted in a global context. Indeed, our identified OTUs restricted to inshore and offshore waters were usually 100% identical to the same morphospecies sequenced in other coastal and oceanic environments, respectively (Supplementary Table S7). Although data on markers less conserved than SSU rDNA are still needed, this suggests ubiquitous dispersal within each kind of environment (Finlay, 2002). However, truly restricted species are also known (for example, the Antarctic *Cymatocylis* spp.; Alder, 1999), so tintinnids may support a model of microbial biogeography where some species are cosmopolitan and others are endemic (Foissner, 2008).

c) Correlation with environmental variables. Environmental factors explained tintinnid β -diversity according to the Mantel tests, although significant results were more frequent for morphospecies than for OTUs (Table 1). Because environmental variables can covary with spatial distance in both the horizontal and vertical dimensions (Martiny *et al.*, 2006), we controlled distance effects using partial Mantel tests, which still indicated significant results for morphospecies, but not for OTUs (Table 1). Compared with morphospecies, the weaker OTU results may be linked to their lower consistency to detect β -diversity patterns in our analyses

(see previous section). This may also explain the unclear relationship between environmental variables and OTU distribution found for choreotrichs and oligotrichs (Doherty *et al.*, 2010; Tamura *et al.*, 2011; Grattepanche *et al.*, 2015).

Ecological processes linked to spatial patterns

The variables that best explained significant relationships between β -diversity and environmental variables in the CMD were proxies for food resources (chlorophyll *a* concentration and microphytoplankton ratio), which showed a higher relevance than hydrographic variables (temperature, salinity and oxygen) or the biomass of potential predators (Table 1). Although we estimated food composition proxies only at the CMD, the general results also support a role of potential prey as β -diversity drivers. The morphospecies associations in each sample group differed in the size of the lorica aperture (that is, oral diameter), considering either the CMD or all samples (Figure 6). Given that lorica oral diameter and prey size are correlated and tintinnids can consume prey as large as about half of their oral diameter (Dolan, 2010; Dolan *et al.*, 2013b), the species that dominated inshore and offshore would consume mainly microplankton and nanoplankton, respectively (Figure 6). Thus, competitive exclusion by dominant species could be linked to the changes in phytoplankton size fraction ratios estimated in the CMD and expected across the shelf as vertical mixing decreases (Figure 1d, Supplementary Figure S2A). Small species from deep waters may even consume bacteria (Figure 6), which would be an alternative resource given the scarcity of phytoplankton below the photic zone (Supplementary Figure S2B; Grattepanche *et al.*, 2015).

Table 1 Simple and partial Mantel tests for the correlations of tintinnid assemblage dissimilarity versus environmental, horizontal or vertical distance (Pearson's coefficient = *R*)

Layer	Simple Mantel test comparison	R	Partial Mantel Test control factor	R	N
All	Morphospecies vs environment ^a	0.31**	Horizontal distance	0.25**	86
			Vertical distance	0.19**	86
	Morphospecies vs horizontal distance	0.36**	Environment ^a	0.30**	86
			Vertical distance	0.36**	86
	Morphospecies vs vertical distance	0.26**	Environment ^a	n.s.	86
			Horizontal distance	0.26**	86
	OTUs vs environment ^a	0.19*	Horizontal distance	n.s.	28
			Vertical distance	0.20*	28
	OTUs vs horizontal distance	0.48**	Environment ^a	0.46**	28
			Vertical distance	0.48**	28
	OTUs vs vertical distance	n.s.	Environment ^a	n.s.	28
			Horizontal distance	n.s.	28
CMD	Morphospecies vs environment ^b	0.35** (<i>r</i> = 0.45**)	Horizontal distance	0.17* (<i>r</i> = 0.34*)	23
	Morphospecies vs horizontal distance	0.34**	Environment ^b	n.s.	23
	OTUs vs environment ^b	0.59** (<i>r</i> = 0.67**)	Horizontal distance	n.s.	7
	OTUs vs horizontal distance	0.59**	Environment ^b	n.s.	7

Bonferroni corrected $P < 0.01$ (**), $P < 0.05$ (*) or $P > 0.05$ (n.s.); number of samples (*N*) exclude tintinnid non-detections. In the chlorophyll maximum depth (CMD), additional BIOENV analyses (Spearman's coefficient = *r*) indicated chlorophyll *a* concentration and microphytoplankton ratio as the subset of environmental variables maximizing significant correlations, both for morphospecies and for OTUs. Two alternative environmental matrices were used:

^aVariables measured in all the samples (temperature, salinity, oxygen concentration and chlorophyll fluorescence).

^bMesozooplankton biomass in the water column and CMD data (temperature, salinity, oxygen concentration, chlorophyll *a* concentration and phytoplankton size fraction ratios).

Figure 6 Proportion of cells by their corresponding lorica oral diameter. Loricae with oral diameters of 41–50, 31–40 and ≤ 30 μm prevailed in the inshore, offshore and deep sample groups, respectively. Size-scaled pictures exemplify size fractions and main taxonomic groups in Figure 2 (from the largest to the smallest: *Stenosemella*, *Amphorides*, *Eutintinnus* and *Salpingella*; loricae > 50 μm excluded).

Consequently, it is possible that tintinnid distribution is structured by the use of different food resources across the shelf. This supports environmental selection, particularly food partitioning, as a mechanism for assemblage structuring in planktonic ciliates (Sitran *et al.*, 2009; Claessen *et al.*, 2010; Wickham *et al.*, 2011). To further explore the processes that structure tintinnid assemblages, we used the distance–decay relationship, which seeks to correlate assemblage dissimilarity with distance between samples (Hanson *et al.*, 2012; Nemergut *et al.*, 2013). Mantel tests indicated that assemblage dissimilarity increased significantly as horizontal distance between samples increased (although CMD results became non-significant when controlling for environmental variables), while it was not significantly affected by vertical distance (Table 1; Supplementary Figure S8). Given that dispersal between neighboring inshore and offshore waters is potentially unlimited, the distance–decay relationship across the shelf supports contemporary environmental selection as more important than evolutionary processes in assemblage structuring (Hanson *et al.*, 2012; Monier *et al.*, 2015). Even if water mixing is expected to transport species across the shelf, environmental selection could prevent establishment (growth) when conditions are non-optimal, and thus limit the successful dispersal (colonization) in the new environment. An additional or alternative process operating on the shelf may be drift (for example, random dispersal), especially in determining the distribution of ecologically similar species within inshore, offshore and deep waters (Dolan *et al.*, 2007, 2009).

Conclusions

Combining DNA sequences with morphological data (by parallel microscope examination and/or via barcoded morphospecies included in reliable databases) is critical to realize the full extent of protist diversity. Contrary to our expectations, OTUs

and morphospecies show a general agreement in assemblage composition, distribution patterns and relationships with environmental variables, with the exception of known biases in HTS quantification. Data integration, however, is crucial because each approach provides unique insights. On one hand, molecular OTUs detected by HTS include hidden diversity, even within well-characterized taxa. These OTUs comprise real species: all the ones observed in the microscope, and also others represented only by few environmental sequences. Furthermore, the distribution of unique OTUs suggests that cryptic species are not distinguished by regular morphological and molecular analyses, unless custom bioinformatics is used (for example, 100% OTU clustering for our tintinnid amplicon). Remaining HTS uncertainties (for example, intra-specific variability and sequencing errors) will be lowered as reference databases become populated with known DNA sequences, thus consolidating HTS usefulness to study, for example, the rare biosphere. On the other hand, morphological characterization helps to understand the processes that structure assemblages, for example, by associating distribution with potential food resources. Our cross-shelf data allow tracking assemblage fluctuations over an environmental gradient and support food partitioning as a mechanism for community assembly in marine plankton. Tintinnids exemplify the necessity of linking data from genotypes (for example, SSU rDNA and more variable markers) and phenotypes (for example, morphology, physiology and ecology) in order to understand the patterns and processes of microbial biogeography.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We are grateful to the crew of the R/V Cape Hatteras and to J Andrade, D Biemesderfer, K Brien, A Gordon, A Liefeld, F Raymond, C Rose and M Tian for assistance during sampling. We also thank the Center for Genome Innovation (University of Connecticut) for technical support for 454 sequencing. This work was supported by the National Science Foundation (Grant OCE1130033 to GBM and OCE-1129734 to LAK), the University of Connecticut and Smith College.

References

- Alder VA. (1999). Tintinninoidea. In: Boltovskoy D (ed), *South Atlantic Zooplankton*. Backhuys Publishers: Leiden, Netherlands, pp 321–384.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic Local Alignment Search Tool. *J Mol Biol* 215: 403–410.

- Bachy C, Dolan JR, Lopez-Garcia P, Deschamps P, Moreira D. (2013). Accuracy of protist diversity assessments: Morphology compared with cloning and direct pyrosequencing of 18s rRNA genes and its regions using the conspicuous tintinnid ciliates as a case study. *ISME J* **7**: 244–255.
- Bachy C, Moreira D, Dolan JR, López-García P. (2014). Seasonal dynamics of free-living tintinnid ciliate communities revealed by environmental sequences from the north-west Mediterranean Sea. *FEMS Microbiol Ecol* **87**: 330–342.
- Bojanić N, Vidjak O, Šolić M, Krstulović N, Brautović I, Matijević S *et al.* (2012). Community structure and seasonal dynamics of tintinnid ciliates in Kaštela Bay (middle Adriatic Sea). *J Plankton Res* **34**: 510–530.
- Boltovskoy D. (1999). *South Atlantic Zooplankton*. Backhuys Publishers: Leiden, Netherlands.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.
- Casteleyn G, Leliaert F, Bockeljaun T, Debeer A-E, Kotaki Y, Rhodes L *et al.* (2010). Limits to gene flow in a cosmopolitan marine planktonic diatom. *PNAS* **107**: 12952–12957.
- Claessens M, Wickham SA, Post AF, Reuter M. (2010). A paradox of the ciliates? High ciliate diversity in a resource-poor environment. *Mar Biol* **157**: 483–494.
- Darriba D, Taboada GL, Doallo R, Posada D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**: 772.
- Doherty M, Tamura M, Costas BA, Ritchie ME, McManus GB, Katz LA. (2010). Ciliate diversity and distribution across an environmental and depth gradient in Long Island Sound, USA. *Environ Microbiol* **12**: 886–898.
- Dolan JR. (2010). Morphology and ecology in tintinnid ciliates of the marine plankton: correlates of lorica dimensions. *Acta Protozool* **49**: 235–344.
- Dolan JR. (2015). Planktonic protists: little bugs pose big problems for biodiversity assessments. *J Plankton Res*; e-pub ahead of print 14 September 2015; doi:10.1093/plankt/fbv079.
- Dolan JR, Pierce RW. (2013). Diversity and distributions of tintinnids. In: Dolan JR, Montagnes DJS, Agatha S, Coats DW, Stoecker DK (eds), *The Biology and Ecology of Tintinnid Ciliates*. Wiley-Blackwell: Oxford, UK, pp 214–243.
- Dolan JR, Ritchie M, 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 JR, Ritchie ME, Tunin-Ley A, 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.
- Dolan JR, Montagnes DJS, Agatha S, Coats DW, Stoecker DK. (2013a). *The Biology and Ecology of Tintinnid Ciliates*. Wiley-Blackwell: Oxford, UK.
- Dolan JR, Landry MR, Ritchie ME. (2013b). The species-rich assemblages of tintinnids (marine planktonic protists) are structured by mouth size. *ISME J* **7**: 1237–1243.
- Edgar RC. (2004). Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Edgar RC. (2010). Search and clustering orders of magnitude faster than blast. *Bioinformatics* **26**: 2460–2461.
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). Uchime improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.
- Egge E, Bittner L, Andersen T, Audic S, de Vargas C, Edvardsen B. (2013). 454 pyrosequencing to describe microbial eukaryotic community composition, diversity and relative abundance: a test for marine haptophytes. *PLoS ONE* **8**: e74371.
- Finlay BJ. (2002). Global dispersal of free-living microbial eukaryote species. *Science* **296**: 1061–1063.
- Foissner W. (2008). Protist diversity and distribution: some basic considerations. *Biodivers Conserv* **17**: 235–242.
- Fontaneto D. (2011). *Biogeography of Microscopic Organisms*. Cambridge University Press: New York, NY, USA.
- Gong J, Dong J, Liu X, Massana R. (2013). Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist* **164**: 369–379.
- Grattepance JD, Santoferrara LF, McManus GB, Katz LA. (2014). Diversity of diversity: conceptual and methodological differences in biodiversity estimates of eukaryotic microbes as compared to bacteria. *Trends Microbiol* **22**: 432–437.
- Grattepance JD, Santoferrara LF, McManus GB, Katz LA. (2015). Distinct assemblage of planktonic ciliates dominates both photic and deep waters on the New England shelf. *Mar Ecol Prog Ser* **526**: 1–9.
- Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.
- Heywood JL, Sieracki ME, Bellows W, Poulton NJ, Stepanauskas R. (2011). Capturing diversity of marine heterotrophic protists: one cell at a time. *ISME J* **5**: 674–684.
- Ishitani Y, Ujiié Y, Takishita K. (2014). Uncovering sibling species in Radiolaria: evidence for ecological partitioning in a marine planktonic protist. *Mol Phylogenet Evol* **78**: 215–222.
- Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, Baerlocher L *et al.* (2011). Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *PNAS* **108**: 13177–13182.
- Li WKW. (2002). Macroecological patterns of phytoplankton in the Northwestern North Atlantic Ocean. *Nature* **419**: 154–157.
- 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* **14**: 813–821.
- Lozupone CA, Knight R. (2008). Species divergence and the measurement of microbial diversity. *FEMS Microbiol Rev* **32**: 557–578.
- Lynch MD, Neufeld J. (2015). Ecology and exploration of the rare biosphere. *Nat Rev Microbiol* **13**: 217–229.
- Magurran A. (2003). *Measuring Biological Diversity*. Wiley-Blackwell: Oxford, UK, p 264.
- Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al.* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**: 102–112.
- Medinger R, Nolte V, Pandey RV, Jost S, Ottenwälder B, Schlötterer C *et al.* (2010). Diversity in a hidden world: potential and limitation of next-generation sequencing

- for surveys of molecular diversity of eukaryotic microorganisms. *Mol Ecol* **19**: 32–40.
- Modigh M, Castaldo S, Saggiomo M, Santarpia I. (2003). Distribution of tintinnid species from 42° N to 43° S through the Indian Ocean. *Hydrobiologia* **503**: 251–262.
- Monier A, Comte J, Babin M, Forest A, Matsuoka A, Lovejoy C. (2015). Oceanographic structure drives the assembly processes of microbial eukaryotic communities. *ISME J* **9**: 990–1002.
- Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF et al. (2013). Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* **77**: 342–356.
- Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R et al. (2014). Vegan: Community Ecology Package. R package version 2.1-41/r2867. Available from <http://R-Forge.R-project.org/projects/vegan/>.
- Pommier T, Canbäck B, Riemann L, Boström KH, Simu K, Lundberg P et al. (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16**: 867–880.
- Reeder J, Knight R. (2009). The 'rare biosphere': a reality check. *Nat Methods* **6**: 636–637.
- Santoferrara LF, Alder VA. (2012). Abundance and diversity of tintinnids (planktonic ciliates) under contrasting levels of productivity in the Argentine Shelf and Drake Passage. *J Sea Res* **71**: 25–30.
- Santoferrara LF, McManus GB, Alder VA. (2013). Utility of genetic markers and morphology for species discrimination within the order Tintinnida (Ciliophora, Spirotrichea). *Protist* **164**: 24–36.
- Santoferrara LF, Grattepanche JD, Katz LA, McManus GB. (2014). Pyrosequencing for assessing diversity of eukaryotic microbes: analysis of data on marine planktonic ciliates and comparison with traditional methods. *Environ Microbiol* **16**: 2752–2763.
- Santoferrara LF, Tian M, Alder VA, McManus GB. (2015). Discrimination of closely related species in tintinnid ciliates: new insights on crypticity and polymorphism in the genus *Helicostomella*. *Protist* **166**: 78–92.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.
- Sitran R, Bergamasco A, Decembrini F, Guglielmo L. (2009). Microzooplankton (tintinnid ciliates) diversity: coastal community structure and driving mechanisms in the southern Tyrrhenian Sea (Western Mediterranean). *J Plankton Res* **31**: 153–170.
- Stamatakis A. (2006). RaxML-vi-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stoeck T, Breiner HW, Filker S, Ostermaier V, Kammerlander B, Sonntag B. (2014). A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. *Environ Microbiol* **16**: 430–444.
- Tamura M, Katz LA, McManus GB. (2011). Distribution and diversity of oligotrich and choreotrich ciliates across an environmental gradient in a large temperate estuary. *Aquat Microb Ecol* **64**: 51–67.
- van der Gast CJ. (2015). Microbial biogeography: the end of the ubiquitous dispersal hypothesis? *Environ Microbiol* **17**: 544–546.
- Weiner A, Aurahs R, Kurasawa A, Kitazato H, Kucera M. (2012). Vertical niche partitioning between cryptic sibling species of a cosmopolitan marine planktonic protist. *Mol Ecol* **21**: 4063–4073.
- Wickham SA, Steinmair U, Kamennaya N. (2011). Ciliate distributions and forcing factors in the Amundsen and Bellingshausen seas (Antarctic). *Aquat Microb Ecol* **62**: 215–230.
- Zinger L, Gobet A, Pommier T. (2012). Two decades of describing the unseen majority of aquatic microbial diversity. *Mol Ecol* **21**: 1878–1896.

Supplementary Information accompanies this paper on The ISME Journal website (<http://www.nature.com/ismej>)