Global and local DNA (meta)barcoding reveal new biogeography patterns in tintinnid ciliates

LUCIANA F. SANTOFERRARA1,2*, EWELINA RUBIN1† AND GEORGE B. MCMANUS1

1DEPARTMENT OF MARINE SCIENCES, UNIVERSITY OF CONNECTICUT, 1680 SHENNECOSSET RD., GROTON, CT, 06340, USA AND 2DEPARTMENT OF ECOLOGY AND EVOLUTIONARY BIOLOGY, UNIVERSITY OF CONNECTICUT, 215 N EAGLEVILLE RD., STORRS, CT, 06269, USA

1PRESENT ADDRESS: GRADUATE SCHOOL OF OCEANOGRAPHY, UNIVERSITY OF RHODE ISLAND 215 S FERRY RD., NARRAGANSETT, RI, 02882 USA,

*CORRESPONDING AUTHOR: luciana.santoferrara@uconn.edu

Received January 15, 2018; editorial decision March 19, 2018; accepted March 21, 2018

Corresponding editor: John Dolan

Tintinnid ciliates are suitable models to study the diversity and biogeography of microbial plankton. In addition to morphological data accumulated over two centuries, most known families and common genera have been linked to DNA sequences in relatively recent barcoding efforts. This backbone of morphologically identified sequences is used here to classify environmental sequences in order to study global and local spatial trends. Analyses of tintinnid SSU rDNA data collected worldwide (about 900 sequences available in NCBI GenBank) and in the northwest Atlantic Ocean (about 500,000 sequences obtained by metabarcoding) support distribution patterns related to salinity, bathymetry and climate/latitude. In addition to the marine-freshwater dichotomy and a pattern of coastal-only taxa known for tintinnids, there is a global trend of phylotypes restricted to brackish or open waters. Local alpha and beta diversity analyses show that assemblage differences among estuarine, coastal and open waters are not significant regarding richness, but are significant in terms of phylogenetic composition. We also confirm spatial restriction of boreal and auroral taxa, and stress that cosmopolitanism cannot be assessed by molecular methods that lump data from potentially endemic and commonly widespread taxa. Heterogeneous diversity, biogeography and phylogenetic resolution within and among tintinnid lineages raise questions about the processes that promote their diversification and determine their spatial distributions.

KEYWORDS: protist; spatial distribution; endemism; cosmopolitanism; brackish-water species.

available online at academic.oup.com/plankt

© The Author(s) 2018. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/plankt/article-abstract/40/3/209/4969366 by University of Connecticut user on 30 May 2018
INTRODUCTION

Tintinnid ciliates are ubiquitous components of microzooplankton, have a mostly marine distribution and are suitable models for exploring the diversity and biogeography of microbial plankton (Dolan et al., 2013). Although these protists usually present a relatively low abundance, they are important trophic links in planktonic food webs and can occasionally consume most of the phytoplankton production in coastal and oceanic systems (McManus and Santoferrara, 2013). Tintinnids are morphologically conspicuous due to their lorica, which has been traditionally used for species identification in taxonomic and ecological studies (e.g. Kofoid and Campbell, 1929). Lorica-based surveys have shown that the structure of tintinnid assemblages changes over geographical scales (Modigh et al., 2003; Thompson and Alder, 2005), vertical profiles (Alder and Boltovskoy, 1993), salinity gradients (Godhantaraman and Uye, 2003; Urrutxutu, 2004) and seasons (Kamiyama and Tsujino, 1996; Modigh and Castaldo, 2002; Bojanic et al., 2012). Abiotic factors such as temperature and salinity, biotic interactions such as phytoplankton grazing (Stoecker et al., 2000), predation by copepods (Dolan and Gallegos, 2001), parasitism by dinoflagellates (Coats et al., 2012) and association with living diatoms (Ambrecht et al., 2017; Vincent et al., 2018), as well as random dispersal (Dolan et al., 2007), influence tintinnid distribution in space and time.

Based on lorica analyses, tintinnids show global biogeographical patterns (Pierce and Turner, 1993; Dolan and Pierce, 2013). A compilation of almost 300 studies in 1 800 locations worldwide has divided tintinnid genera into five biogeographical categories that match the distribution patterns known for many planktonic organisms (Dolan and Pierce, 2013): cosmopolitan (from Arctic to Antarctic, not restricted to nearshore waters), neritic (from Arctic to Antarctic, but restricted to nearshore waters), warm-temperate (both coastal and open waters, but not at polar and sub-polar latitudes), boreal (both coastal and open waters, Arctic and Subarctic) and austral (both coastal and open waters, Antarctic and Subantarctic).

In the past 15 years, tintinnid molecular information, mostly based on complete or partial sequences of the small-subunit ribosomal RNA gene (SSU rDNA), has been collected by two main approaches: first, known isolates are the source for morphologically identified sequences that can be used as barcodes; then, this reference can be used to classify environmental sequences, formerly by clone libraries and more recently by metabarcoding (Santoferrara et al., 2016a). Molecular data have confirmed that several lorica-based families and genera are invalid (Snoeyenbos-West et al., 2002; Agatha and Strüder-Kypke, 2013; Santoferrara et al., 2017) and that cryptic and polymorphic species exist (Bachy et al., 2012; Santoferrara et al., 2013, 2015). Although lorica- and DNA-based tintinnid assemblages agree reasonably well at discrete sites (Bachy et al., 2013) and across local environmental gradients (Santoferrara et al., 2016b), global biogeography patterns have not been tested molecularly.

We compiled all of the tintinnid SSU rDNA sequences available in NCBI GenBank, including both morphologically identified and environmental sequences. Almost 900 sequences were retrieved, curated, grouped into phylotypes (based on similarity and phylogenetic relationships) and linked with metadata from 50 corresponding publications. Our first aim was to check for global spatial patterns potentially unappreciated in individual molecular studies, or in the even richer body of morphological information. The results of this meta-analysis led us to two hypotheses that we tested locally by metabarcoding in estuarine, coastal and open waters of the northwest Atlantic Ocean: (i) estuarine waters present phylotypes not detected in neighboring marine waters; and (ii) open waters contain phylotypes not found in adjacent coastal waters.

METHOD

Global barcoding data

All of the tintinnid SSU rDNA sequences in NCBI GenBank were retrieved (last updated on 15 May 2017). This includes both morphologically identified and environmental sequences (the latter, mostly from Sanger-sequenced clone libraries). Retrieval of sequences and associated metadata was done as a part of the EukRef initiative (del Campo et al., 2018). A set of reliably identified sequences were the seed to iteratively retrieve all GenBank sequences with >80% similarity, using the BLASTn algorithm (Camacho et al., 2009) against the NCBI non-redundant/nucleotide collection. Sequences shorter than 500 bp (less reliable for phylogenetic analysis; e.g. Dunthorn et al., 2014), chimeras detected with UCHIME (Edgar et al., 2011) and one dataset with known inconsistencies (GenBank accession numbers AB640624 to AB640682) were excluded, resulting in 870 sequences retained. One sequence obtained here was added (GenBank accession number MG719774; Supplementary Data).

To avoid redundancy within studies, replicated sequences from the same sampling site were removed, resulting in a database of 675 sequences. Each sequence was paired with its sampling site and categorized by environment. Environmental information was retrieved from the corresponding GenBank entry or original publication. If salinity or bathymetry data were unavailable, approximations
were made based on the coordinates of the sampling sites plotted on the salinity map of the World Ocean Atlas (Boyer et al., 2013) or on an ocean relief map (Google Earth). Environmental categories were based on (i) salinity: freshwater (<0.5); brackish (0.5–30; this includes mostly estuaries, and some coastal lagoons and inlets where salinity may be >30 during periods of low freshwater inputs) and marine (>30); (ii) bathymetry (only for marine environments): coastal (<50 m), inner shelf and open waters (>50 m, outer shelf, slope and oceanic); deep coastal sites (reaching >50 m depth at <2 km from shore) were considered separately; and (iii) climate/latitude: warm-temperate, austral (Antarctic and Subantarctic) and boreal (Arctic and Subarctic).

Sequences were grouped into phylotypes based on both pairwise p-distances (estimated in MEGA; Tamura et al., 2011a) and phylogenetic relationships (based on a preliminary RaxML tree built as described at the end of this section, but using only 100 bootstraps; results not shown). Groups of sequences more than 99% similar and phylogenetically cohesive were lumped into a phylotype (note that although these criteria provide the best approximation for species delimitation based on long SSU rDNA sequences, one phylotype does not necessarily mean one species; Santoferrara et al., 2013). No evident spatial pattern was masked by this grouping strategy. For each clustered phylotype, one representative was selected (priority: well-documented isolate > minimally documented isolate > environmental), resulting in 126 sequences (average length = 1573 bp, standard deviation = 215 bp). The number of phylotype detections per environment was recorded, excluding replicated detections from a same study and site. The final matrix includes 351 sequences from 150 sites and 50 published studies (Fig. 1A, Table SI).

The 126 representative phylotypes constitute our reference dataset for further analyses. This dataset was combined with sequences of Choreotrichida Small & Lynn 1985 as outgroup and aligned with MAFFT v. 7 (Katoh and Standley, 2013). Ambiguous positions were removed with the guidance of Gblocks v. 0.91b under default parameters (Castresana, 2000). Maximum likelihood inference was done with RaxML v. 8.3.17 (Stamatakis, 2014); the best-known tree was inferred out of 200 initial trees, and node support was estimated after 10,000 bootstraps. Bayesian inference was done with MrBayes v. 3.2.1 (Ronquist et al., 2012). Five million generations were run and trees were sampled each 1000 cycles. The initial 1000 trees were discarded as burn-in, and the remaining 4000 trees were used to estimate the Bayesian posterior probabilities. For each analysis, the GTR model with a Γ model of rate heterogeneity and a proportion of invariable sites was used, as previously identified with MrModeltest v. 2 (Nylander, 2004) under the Akaike Information Criterion.

**Metabarcoding in the northwest Atlantic and adjacent estuarine waters**

We performed metabarcoding for 51 samples collected in estuarine, coastal and open waters (Fig. 1B; for sampling and methodological details, see Table SII). Coastal and open waters off New England (as delimited by the 50-m isobath) were sampled at two to four depths with Niskin bottles mounted on a CTD rosette on board the R.V. Hatteras in summer 2012 and the R.V. Connecticut in summer 2015. Estuarine waters were sampled at the surface from a dock in the mouth of the Poquonnock estuary on Long Island Sound (spring 2015 to winter 2016) and from a small boat on the tidal part of the Thames River, Connecticut (summer 2016). For each sample, 1–3.4 L of water were concentrated on 3- or 10-μm polycarbonate filters, which were stored in buffer until DNA extraction using phenol/chloroform or commercial kits (Table SII).

A partial V2–V3 region of SSU rDNA (about 308 bp long) was PCR-amplified with primers specific for tintinnids and related ciliates (Tamura et al., 2011b) and adapted for multiplexed high-throughput sequencing. The 2012 tintinnid sequences were obtained with Roche 454 ( Branford, CT) in a previous study (Santoferrara et al., 2016b). The remaining samples were newly amplified with High Fidelity Phusion polymerase (Thermo Scientific, Waltham, MA) under the following conditions: 98°C for 30 s, 30 cycles at 98°C for 10 s, 58°C for 20 s and 72°C for 40 s, and 10 min at 72°C. Three PCR products per sample were pooled and sequenced with MiSeq (illumina, San Diego, CA), using a dual-index strategy (Kozich et al., 2013; Nelson et al., 2014).

Sequence demultiplexing and quality filtering were done in QIIME v. 1.9 (Caporaso et al., 2010), as optimized for tintinnids and related ciliates (Grattepanche et al., 2014; Santoferrara et al., 2014, 2016b). Chimeras identified with the UCHIME de novo algorithm (Edgar et al., 2011) and the rarest sequences (with five or fewer reads) were removed. The remaining sequences were clustered de novo with UCLUST (Edgar, 2010) using a 100% similarity cut-off. This approach maximizes the detection of short sequence variants, thus offering the best approximation for species-level analysis with our target DNA region (Santoferrara et al., 2014, 2016b). Variants were assigned taxonomically with BLASTN (Altschul et al., 1990) against our reference dataset of 126 representative phylotypes.

In total, we retained 518,492 high-quality tintinnid reads (Table SII). Alpha and beta diversity analyses were done as implemented in QIIME v. 1.9 (Caporaso et al., 2010).
et al., 2010). To avoid heterogeneity due to sequencing depth, the dataset was sub-sampled randomly, retaining 2000 sequences per sample. Alpha diversity was estimated as the observed number of variants, the Chao1 estimator of total variant richness and the PD Whole Tree metric (that considers both the number of variants and their phylogenetic relationship). These parameters were tested for statistical difference ($\alpha = 0.05$) between sample groups with a non-parametric $t$-test using 1000 Monte Carlo permutations and the Bonferroni correction. For beta diversity analyses, unweighted and weighted UniFrac distance matrices were generated, then used for principal coordinates analyses. The significance ($\alpha = 0.05$) of assemblage differences among sample groups was tested with PERMANOVA and PERMDISP, two non-parametric multivariate analyses of variance that are not affected by unbalanced sampling design (Anderson, 2001). To compare assemblage composition among estuarine, coastal and open waters, samples were pooled by environment and standardized to the minimum sequencing depth (ca. 50,000 reads per sample group).

RESULTS AND DISCUSSION

Tintinnid phylogenetic structure

Tintinnid SSU rDNA data collected worldwide and available in GenBank are represented by 126 phylotypes (groups of sequences phylogenetically cohesive and 99% similar). Phylogenetic inference (Fig. 2) showed Tintinnidiidae, Tintinnidae, Eutintinnidae and Favellidae placed sequentially (only disrupted by three unidentified phylotypes), and the remaining families arranged in poorly supported nodes, in agreement with previous reports (Agatha and Strüder-Kypke, 2013; Santoferrara et al., 2017).

Because they were generally clustered within known families and genera, most phylotypes derived exclusively from environmental sequences probably correspond to known lineages for which identified specimens have not yet been sequenced. This barcoding information is lacking for 1 out of the 14 families and 37 out of the 76 genera included in a recent review of tintinnid taxa (Santoferrara et al., 2017; plus the recently added Dartinnimus Smith & Santoferrara 2018). The yet un-sequenced Nolachusiidae may correspond to the environmental clade placed (with low support) between Tintinnidae and Eutintinnidae (Fig. 2), as expected based on cell features (Agatha and Strüder-Kypke, 2013). As discussed below, the putative Nolachusiidae sequences were detected in brackish waters, in agreement with the existing reports of Nolachusiis, the only genus in this family (Sniezek et al., 1991; Snyder and Brownlee, 1991). Between the putative Nolachusiidae and the known Eutintinnidae, however, there is a distinct phylotype (represented by GenBank sequence EU333101) from open waters that has unconfirmed affiliation (Fig. 2).

Although almost half of the known tintinnid genera have not been sequenced based on morphologically
Fig. 2. Tintinnid phylogenetic tree (right) and phylotype detection per environment (left) according to the existing GenBank data (additional details in Fig. S1). Environments are classified based on salinity (S), bathymetry (B) and climate/latitude (C/L). Marine reports include detections in extreme environments.

identified isolates yet, relatively few phylotypes with genus-level divergence derived exclusively from environmental sequences (Fig. 2). Apart from the mentioned *Nolaclusilis*, the environmental clade placed within Tintinnidiidae could correspond to *Membranicola*, the only known genus in this family that remains to be bar-coded. Of the other un-sequenced genera, however, only 10 pass the condition of being reported at least four times by two different authors (Dolan and Pierce, 2013). This suggests that many of the genera not yet
Global distribution patterns

Comparing the phylogenetic tree and the spatial distribution of phylotypes shows that none of the tintinnid families is truly restricted by environment (Fig. 2; except Tintinnidae, which appears to be restricted to marine waters, but see next section). Instead, phylotypes from different environments are interspersed in the tree. Phylo­type detection based on the compiled data (50 publications and 150 sampling locations; Fig. 1A; Table SI) covers relatively few areas of the world, biased towards a few sites with concentrated sampling (neritic waters off China and Northeast USA, Mediterranean Sea). Because of undersampling, a conclusive molecular evaluation of tintinnid biogeography is presently not possible, but the information we compiled was complete enough to show global distribution patterns related to salinity, bathymetry and climate/latitude (this section) and to generate novel hypotheses that we tested locally by metabarcoding (next section).

Salinity

Most phylotypes were reported in marine waters (79%) and fewer in freshwater (9%), as expected for tintinnids (Fig. 2). These ciliates apparently originated from a marine ancestor (Agatha and Strüder-Kypke, 2013) and transitioned towards freshwater only rarely and recently (Bachy et al., 2012). Most freshwater taxa correspond to a single family (Tintinnidiidae), although one of its clades (Tintinnidium clade I) includes only sequences from brackish and marine waters so far. Three other scattered phylotypes were detected in freshwater (two Tintinnopsis species and one unidentified Undelliidae).

Some phylotypes are apparently restricted to brackish waters (Fig. 2). In agreement with our data, species in Nolachusis and Daritintinnus have been described under this water regime (Sniezek et al., 1991; Snyder and Brownlee, 1991; Smith et al., 2018). Other reports for these taxa are rare, but also come from brackish waters in Chesapeake Bay (Stoecker et al., 2000), Rio de la Plata estuary (Kogan, 2005) and the Black Sea (Gavrilova and Dolan, 2007). Surprisingly, however, some phylotypes related to Daritintinnus were detected in deep anoxic or hypersaline waters (Stoeck et al., 2006; Edgcomb et al., 2009). Detection of tintinnid DNA in these and other extreme environments (some Tintinnidae detected in the same and additional studies, for example in hydrothermal fields; López-García et al., 2007; Edgcomb et al., 2011) may relate to their capacity to form cysts, which are known to survive adverse conditions such as anoxia and high sulfide concentration (Kamiyama, 2013).

Bathymetry

Most marine phylotypes were reported in either coastal (28%) or open waters (30%), with only 16% detected in both kinds of environments (Fig. 2; the remaining phylotypes are from deep coastal sites and are considered separately below). The differentiation between coastal and open water phylotypes agrees with previous studies that have focused on tintinnids across the gradient from coast to ocean. On wide or relatively wide shelves of the southwest Atlantic, East China Sea and northwest Atlantic (about 600, 500 and 150 km wide, respectively), tintinnid assemblages change considerably between the 50 and 100 m isobaths (Santoferrara and Alder, 2012; Li et al., 2016; Santoferrara et al., 2016b). Accordingly, there are genera differentially distributed in either coastal (e.g. Favella, Helicostomella) or open waters (e.g. Parandella, Xystonella). While this agrees with the well-known differentiation of exclusively neritic taxa worldwide (Pierce and Turner, 1993; Dolan and Pierce, 2013), a pattern of tintinnids that are exclusive to open waters has not been recognized in previous global surveys.

The coastal versus open waters pattern is not absolute, as the exchange of taxa due to oceanographic phenomena (e.g. surface or deep currents, upwelling and mesoscale eddies) can result in occasional dispersal (but usually not colonization) outside the normal range (Balech, 1972; Bolotovsky and Alder, 1992; Kato and Taniguchi, 1993; Kim et al., 2012). Another factor that influences the bathymetric pattern is geomorphology. In deep coastal sites such as the much-sampled Bay of Villefranche, where the continental shelf drops abruptly (<2 km from shore) and the influence of open waters is constant, both coastal and open water phylotypes are commonly detected (Fig. 2; Bachy et al., 2012, 2013), in agreement with morphology-based studies there (Dolan, 2017) and in other narrow shelves of the Mediterranean (Sitran et al., 2007, 2009).

Climate/latitude

Most phylotypes represented in GenBank were obtained in warm-temperate environments (Fig. 2). However, it is clear that phylotypes corresponding to known austral or boreal genera were not detected outside their expected range. Phylotypes corresponding to Cymatocylis and Laackmanniella were detected only in Antarctica (Picquet et al., 2008;
Kim et al., 2013). Pychoecylis and Parafaviella were detected in the Artic and towards temperate waters, but always in the northern hemisphere (Terrado et al., 2012; Santoferrara et al., 2016b). This implies that the restriction of these austral or boreal genera known from a large body of microscopy reports (Pierce and Turner, 1993; Alder, 1999; Dolan et al., 2012, 2017; Dolan and Pierce, 2013) is real and not due to loricax taxonomic limitations.

Recent metabarcoding surveys focused on tinnitids in the northwest Atlantic or the Mediterranean neither detected austral genera (Bachy et al., 2013; Santoferrara et al., 2016b). Based on ciliate metabarcoding data from the Tara Oceans expedition, instead, it has been suggested that Laeckmanniella is not restricted to austral waters, but that it is a cosmopolitan genus (Gimmler et al., 2016). Despite the relevance of this circumboreal dataset, the short region evaluated (V9) and the clustering threshold applied (97%) very likely resulted in a methodological artifact, as the approach used by Gimmler et al. (2016) cannot distinguish Laeckmanniella from several widely distributed taxa (e.g. Codonellopsis, Stramenella and some Tintinnopsis; Fig. S2).

At the species level, we found discrepancy regarding Amphorellopsis quinqeualata (Fig. 2). Based on loricar morphology, this species is believed to be restricted to the Southern Ocean (Dolan et al., 2012), but its sequence presents a 99.8–99.9% similarity with relatively long environmental sequences (ca. 1340 bp) from the Mediterranean (Bachy et al., 2013, 2014). The fact that neither loricar (due to crypticity, polymorphism and other taxonomic issues) nor SSU rDNA sequences (given the lack of a universal interspecific divergence threshold) are definitive for species delineation means that some distributional hypotheses cannot presently be tested. In part for these reasons, global analyses of distribution have not focused on species, but on genera (Dolan and Pierce, 2013; this study). At this level, morphological and molecular distributions agree genus by genus, except for the putative cosmopolitans that we cannot assess molecularly with our relatively small SSU rDNA dataset.

Local metabarcoding analysis

Alpha and beta diversity

Tintinnid assemblage composition varied in estuarine, coastal and open waters of the northwest Atlantic (Fig. 3). The three sample groups differed significantly based on both unweighted (presence/non-detection) and weighted (by relative sequence abundance) UniFrac distance matrices of tinnitid variants, except when the unweighted metric was tested with PERMDISP (Fig. 3). The weaker signal of the non-quantitative analysis suggests that assemblage differentiation is less substantial in terms of presence/non-detection than in terms of relative abundances (Lozupone et al., 2007). In other words, this supports the idea that many taxa of planktonic protists undergo occasional expatriation (resulting in rare specimens detected outside their habitat), but do not thrive in the new environment (i.e. effective dispersal is much lower than the potentially unlimited dispersal of microbes; Weisse, 2008).

Contrary to assemblage composition, metrics of alpha diversity did not differ significantly among sample groups (Fig. 4). Although these data have only relative value (for example, because bioinformatic methods designed to maximize variant detection may also capture some intraspecific or erroneous variants; Santoferrara and McManus, 2017), our results indicate that diversity is not lower in estuaries as compared to adjacent environments. Previous molecular studies of planktonic ciliates (choereticrichs and oligotrichs) have shown higher alpha diversity where the outflow plume of the Connecticut River reaches Long Island Sound, compared to purely riverine and marine neighboring waters (Doherty et al., 2010; Tamura et al., 2011b). Tintinnid diversity has also shown an indirect relationship with salinity, as well as higher values than expected when contrasted to typical marine environments, in the estuarine waters of Chesapeake Bay (Dolan and Gallegos, 2001). These trends appear to contradict the long-held notion that biodiversity is minimum in brackish waters as compared to freshwater and marine regimes (Remane, 1934). This idea, however, was based only on macrozoobenthos. Instead, the diversity of bacteria and protists is actually maximum in brackish waters (Telesh et al., 2013; Cepleti et al., 2017), or at least not significantly different (Herlemann et al., 2011; Hu et al., 2016), in relation to adjacent environments.

Comparison of metabarcoding variants and reference phylotypes

All variants detected by metabarcoding could be assigned to one of the reference phylotypes obtained from GenBank. Out of the 126 reference phylotypes, 52 were detected in the northwest Atlantic dataset with a BLASTN identity higher than 99%. Thus, about 40% of the phylotypes obtained in different parts of the world have a closely related, if not an identical, match at the local scale studied here (note that this analysis excluded matches to five phylotypes that were previously reported only in this area).

The prevalent phylotypes were detected exclusively or almost exclusively in only one of the three environments surveyed (Fig. 5), which explains the assemblage differentiation among sample groups (Fig. 3). Each environment contained characteristic phylotypes: Dartintinnum alderae, the putative Noctiluopsis, Tintinnidium mucicola, Favella panamensis and Tintinnidium baleich in estuarine...
Tintinnid assemblages differed significantly in the estuarine, coastal and open waters assessed by metabarcoding. The separation of sample groups is based on unweighted (A) and weighted (B) UniFrac distance.

Fig. 3. Tintinnid assemblages differed significantly in the estuarine, coastal and open waters assessed by metabarcoding. The separation of sample groups is based on unweighted (A) and weighted (B) UniFrac distance.

waters; *Stenosemella pacifica* and *Salpingacantha undata* in coastal waters; and *Stenstrupiella stenstrupii*, *Eutintinnus permimus*, *Salpingacantha unguiculata* and *Dictyocysta lepida* in open waters. The exception was the most abundant phylotype, which comprised 3%, 46% and 51% reads in estuarine, coastal or open waters, respectively (Fig. 5), and corresponds to an uncultured organism (GenBank sequence KJ758267) related to *Salpingella*. Recent evidence suggests that this genus is more abundant and diverse than previously thought (Santoferrara et al., 2016b), and that some of its species form symbiotic relationships with diatoms (Vincent et al., 2018), thus uncovering an unexpected ecological importance of *Salpingella*.

Testing biogeography hypotheses

The partition of our metabarcoding reads into estuarine, coastal and open waters is used here to test hypotheses derived from the compilation of GenBank data. Contrary to our GenBank survey, which is likely influenced by the limited sampling power of Sanger-sequenced isolates and clone libraries, metabarcoding allows for detection of rare phylotypes and standardization of sequencing depth among samples. In addition to re-confirming that boreal (but not austral) phylotypes are detected in the northern hemisphere waters investigated (Fig. S3), our results support the hypothesized restriction of some phylotypes to either estuarine or open waters (Table I).

Estuarine lineages (at least some species in *Daritinnus*, *Eutintinnus*, *Tintinnopsis* and the putative *Nolaclusilis*) were not detected in adjacent marine waters (their non-detection...
in freshwater, though, remains suggested only by the GenBank data as we did not investigate this kind of environment by metabarcoding). While the marine–freshwater dichotomy is one of the most well-known biogeographic patterns for aquatic microbes (Lozupone and Knight, 2007; Logares et al., 2009; Forster et al., 2012), restriction to brackish or estuarine waters has only recently been established. For example, assemblages of both bacteria and protists differ between brackish waters and adjacent marine or limnic waters in places such as the Baltic Sea and estuaries of North America, based on SSU rDNA (Grump et al., 2004; Herlemann et al., 2011; Hu et al., 2016), metagenomics (Dupont et al., 2014; Hugerth et al., 2015) and patterns of gene expression (Hewson et al., 2014; Celepli et al., 2017). For bacteria, some of these studies suggest that lineages specifically adapted to brackish waters have a long evolutionary history and are dispersed to similar environments globally (Hugerth et al., 2015). Experimentation on salinity tolerance in planktonic ciliates provides evidence for genetically fixed preferences: isolates of the oligohymenophorean Cyclorella glaucoma from freshwater, brackish and hypersaline waters were found to tolerate broad salinity ranges, but grow better at the original salinity (Finlay et al., 2006). The ecophysiological basis for brackish-water restriction in tintinnid taxa remains unexplored, with salinity as one of the possible (and clearly understudied) determining factors. Yet, other variables that correlate with the salinity gradient in land-margin environments (e.g. enrichment of prey that depends on terrestrially derived nutrients) could influence the observed pattern. For example, the genus Rhizodorus has a neritic distribution but it is usually more abundant in estuaries and lagoons, apparently not associated to salinity but to eutrophication (Saccà and Giafrè, 2013).

The open water lineages (e.g. Xystonella) were usually not detected in the adjacent coast (Table I; Fig. S3). This agrees with data for some of the most abundant ciliates in the same area (Grattepanche et al., 2015). It is also consistent with the differentiation of neritic and oceanic taxa in many groups of planktonic protists, which is typically attributed to the contrasting trophic features of these realms (Forster et al., 2012). In our study, exceptions to this trend included some phylotypes clearly prevalent in open waters that presented few reads in one estuarine sample from December 2013, coincident with a salinity value above the typical brackish range (sample G5; Tables I and SII). This was probably due to an episodic influx of adjacent marine water, as reported at the mouth of another estuary in northeast USA, where sporadic detection of oceanic tintinnids was attributed to changes in surface current circulation into

Fig. 5. The most abundant phylotypes in the estuarine, coastal and open waters assessed by metabarcoding.

Table I: Phylotypes reported only in brackish or open waters on GenBank that were detected mostly in estuarine or open waters of the northwest Atlantic by metabarcoding, respectively

<table>
<thead>
<tr>
<th>Phylotype</th>
<th>% estuarine</th>
<th>% coastal</th>
<th>% open water</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brackish</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HQ394065_Nolacaulis sp.?</td>
<td>99</td>
<td></td>
<td>n.d.</td>
</tr>
<tr>
<td>MF039886_Dartinittimus alderae</td>
<td>100</td>
<td></td>
<td>n.d.</td>
</tr>
<tr>
<td>AY180046_Dartinittinus sp.?</td>
<td>100</td>
<td></td>
<td>n.d.</td>
</tr>
<tr>
<td>AF399170_Eutinittimus sp. SW-2002</td>
<td>100</td>
<td></td>
<td>n.d.</td>
</tr>
<tr>
<td>JN634185_Tintinnopsis major</td>
<td>100</td>
<td></td>
<td>n.d.</td>
</tr>
<tr>
<td>Open water</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JG924858_Amphorellis quinquedenta</td>
<td>1*</td>
<td>1</td>
<td>98</td>
</tr>
<tr>
<td>EU399536_Salpingella sp. SK-2008</td>
<td>22*</td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>KT792933_Xystonella longicauda</td>
<td>n.d.</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>JX101863_Protorhabdorina curta</td>
<td>17*</td>
<td></td>
<td>83</td>
</tr>
<tr>
<td>KY290319_Epyrocylis undella</td>
<td>17*</td>
<td></td>
<td>83</td>
</tr>
<tr>
<td>KY290315_Ascamperellia acuta</td>
<td>24*</td>
<td></td>
<td>76</td>
</tr>
<tr>
<td>KY290321_Psychoclycis minor</td>
<td>n.d.</td>
<td></td>
<td>100</td>
</tr>
<tr>
<td>MG719774_Acanthostomella sp.</td>
<td>15*</td>
<td></td>
<td>85</td>
</tr>
<tr>
<td>KY290330_Climacocystis scalaroides</td>
<td>22*</td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>KY290327_Parundella aculeata</td>
<td>1*</td>
<td></td>
<td>98</td>
</tr>
</tbody>
</table>

The asterisk denotes proportions of open water phylotypes in one estuarine sample (G5), probably due to an episodic influx of adjacent marine water (see text). n.d., non-detection.
the adjacent shelf (Sanders, 1987). It is well known, for example, that eddies and meanders of the Gulf Stream occasionally bring oceanic water masses into the area (Chang and Dickey, 2001). These minor temporal variations do not change the established distribution patterns considerably, but exemplify sporadic expatriations known for tintinnids and other planktonic organisms (e.g. Bolotovskoy and Alder, 1992).

Restricted distributions may also relate to parameters not evaluated here. For instance, Metacylis angulata has never been recorded outside estuarine and neritic waters of the northeast USA, but the reasons for its apparent absence in other coastal areas of the world are unclear (Pierce, 1996; Santoferrara et al., 2017). Limited geographical distribution in some species, however, does not mean inexistence of genuine cosmopolitans. Species with identical lorica and rDNA sequences have been detected in, for example, the Atlantic and the Pacific oceans (Santoferrara et al., 2016b; Vincent et al., 2018). Still, a better understanding of tintinnid distribution may arise once morphospecies are studied with more variable and/or non-neutral genetic markers, especially in the current—omics era.

**CONCLUSIONS**

Our understanding of tintinnid phylogenetic structure remains incomplete despite inferences based on increased taxonomic sampling (including clades currently represented by environmental sequences only; this study) or character sampling (including three regions of the rDNA operon; Santoferrara et al., 2017). There are at least two challenges in resolving tintinnid phylogenetic relationships: one, to chase taxa potentially unrepresented in the tree (e.g. by sampling heavily unexplored areas or environments, such as the Southern Hemisphere or the deep ocean), and two, to incorporate new tools and features (e.g. phylogenomics, finer morphological and ecophysiological differentiation).

Tintinnids restricted by salinity, bathymetry or climate/latitude are not phylogenetically or morphologically cohesive. In addition to known marine versus freshwater and coastal-only trends, there are phylotypes constrained to brackish or open waters. Endemism of boreal and austral taxa is confirmed molecularly, while cosmopolitanism remains untested due to limitations in genetic markers and spatial reach. Our local data also show that estuarine waters are not less diverse than adjacent marine environments, and that tintinnid taxa restricted to estuaries are promising targets for exploring adaptations to cope with frequent salinity changes.

While biogeography patterns are relatively well known for tintinnids, the mechanisms that promote their diversification and determine their spatial distributions are not. Local tintinnid assemblages are thought to be structured by contemporary environmental selection (Sitran et al., 2009; Santoferrara et al., 2016b) and random dispersal (Dolan et al., 2007). However, their global diversity and biogeography have been shaped by various evolutionary processes, as suggested by the phylogenetic structure and spatial distribution of tintinnid lineages. For example, we speculate that the austral versus boreal isolation of the closely related Cymatocylis and Psychoclysis is an example of allopatric diversification (Abellán and Riberas, 2017). These and several other morphologically distinct taxa inferred as polytomies in the tintinnid phylogenetic tree may reflect rapid radiation and/or extinctions (as observed, for example, in spathidiid ciliates; Rajter and Vďačný, 2016), thus raising questions about when and why these events happened.

**SUPPLEMENTARY DATA**

Supplementary data can be found online at *Journal of Plankton Research* online.

**ACKNOWLEDGEMENTS**

We thank the crew and participants of the R.V. Hatteras and R.V. Connecticut cruises for assistance with sampling off New England, and S. Smith, R. Cole and L. Sanchez (University of Connecticut, USA) for help with sampling in the Thames River and Long Island Sound. We also thank L. Katz and J.D. Grattepanche (Smith College, USA) for help with sampling planning and preparation of the 2012 DNA library. The Center for Genome Innovation of the University of Connecticut, in particular B. Reese and L. Lu, provided technical support for sequencing. The Computational Biology Core Facility of the University of Connecticut provided resources for phylogenetic analyses.

**FUNDING**

The National Science Foundation (Grant OCE1435515) and the University of Connecticut, USA.

**DATA ARCHIVING**

The tintinnid reference database is publicly available as part of EukRef-Ciliophora on www.eukref.org. One isolate sequence was deposited in NCBI GenBank (accession number MG719774). Metabarcoding data were
REFERENCES


