7-1-2020

High Diversity of Testate Amoebae (Amoebozoa, Arcellinida) Detected by HTS Analyses in a New England Fen using Newly Designed Taxon-specific Primers

Alistaire Ruggiero
Smith College

Jean David Grattepanche
Smith College

Agnes K.M. Weiner
Smith College

Laura A. Katz
Smith College, lkatz@smith.edu

High diversity of testate amoebae (Amoebozoa, Arcellinida) detected by HTS analyses in a New England fen using newly-designed taxon-specific primers

Running header: Arcellinida diversity in a New England fen

Alistaire Ruggiero [a], Jean-David Grattepanche [a], Agnes K. M. Weiner [a], Laura A. Katz [a,b]

[a] Smith College, Department of Biological Science, Northampton, Massachusetts, USA

[b] University of Massachusetts Amherst, Program in Organismic and Evolutionary Biology, Amherst, Massachusetts, USA

Correspondence:
L. A. Katz, Dept Biological Sciences, 44 College Lane, Northampton MA, 01060, USA
Telephone number: +1 413-585-3815, FAX number +1 413-585-3786; email: lkatz@smith.edu

ABSTRACT:

Testate (shell-building) amoebae, such as the Arcellinida (Amoebozoa), are useful bioindicators for climate change. Though past work has relied on morphological analyses to characterize Arcellinida diversity, genetic analyses revealed the presence of multiple cryptic species underlying morphospecies. Here, we design and deploy Arcellinida-specific primers for the SSU-rDNA gene to assess the community composition on the molecular level in two sample sets from a New England fen: 1) 36 cm horizontal transects and vertical cores; and 2) 26 m horizontal transects fractioned into four size classes (2-10 μm, 10-35 μm, 35-100 μm, 100-300 μm). Analyses of these data show: 1) a considerable genetic diversity within Arcellinida, much of which comes from morphospecies lacking sequences on GenBank; 2) communities characterized by DNA (i.e. active + quiescent) are distinct from those characterized by RNA (i.e. active, indicator of biomass); 3) active communities on the surface tend to be more similar to one another than to core communities, despite considerable heterogeneity; and 4) analyses of communities fractioned by size find some lineages (OTUs) that are abundant in disjunct size categories, suggesting possible life history stages. Together these data demonstrate the potential of these primers to elucidate the diversity of Arcellinida communities in diverse habitats.

Keywords: Arcellinida; testate amoebae; primer design; amplicon sequencing; phylogeography.

Testate amoebae of the order Arcellinida (Amoebozoa) are unicellular eukaryotes that construct tests (shells) around their cells. They are commonly found in wetland environments worldwide, such as bogs, fens or lakes, where their tests are preserved in the peat or sediment after the death of the organism (Bosak et al. 2011, Lahr et al. 2015, Kosakyan et al. 2016, Porfirio-Sousa et al. 2017, Charman 1999, Mitchell et al. 2008). Estimates of their abundance in bogs range between 10^3 – 10^4 individuals per gram dry weight of *Sphagnum* (Wilkinson 2008, Mitchell et al. 2008) and they are considered top-predators as they consume bacteria, fungi, microalgae and micro-metazoa (Yeates and Foissner 1995, Creevy et al. 2018).

Testate amoebae are widely used as bioindicators of changing conditions in peatlands, as they are the dominant group of microbial eukaryotes on *Sphagnum* (Gilbert et al. 1998) and are sensitive to environmental conditions. As peatlands contain roughly 30% of global soil carbon, testate amoebae are useful for determining peatland health and the consequential changes in carbon stocks (Turner and Swindles 2012, Yu et al. 2011). Testate amoebae diversity varies in response to changing abiotic factors, such as moisture and pH (Turner and Swindles 2012, Booth et al. 2008, Bobrov et al. 2004, Tolonen et al. 1992, Tolonen et al. 1994). For example, *Lesquereusia* spp., *Planocarina carinata*, and large *Difflugia* spp. are restricted to wet environments and have a higher pH optimum, while the opposite is true for *Hyalosphenia subflava* and *Nebela tincta* (Booth 2001). In addition, testate amoebae are sensitive to environmental pollution, such as road-salt or acid mine drainage and can therefore be used to trace potential contamination of ecosystems (e.g. Patterson et al. 2013, Roe and Patterson 2014).

In order to use Arcellinida as bioindicators, knowing their large- and small-scale distribution patterns, as well as their habitat requirements is of considerable importance. A longstanding debate in protistology has questioned whether species have a cosmopolitan or geographically limited distribution (e.g. Finlay et al. 1999, Finlay and Clarke 1999, Finlay 2002, Finlay et al.

2001, Foissner 1999). Detailed biogeographic studies on the level of morphospecies and genotypes in the last years revealed many examples of restricted distribution patterns, likely due to local adaptations to environmental factors (e.g. Smith et al. 2008, Singer et al. 2019, Heger et al. 2013). In the case of Arcellinida in *Sphagnum* bogs, access to the materials that they use to build their tests may be one of the constraints that determines their micro-distribution (Heger et al. 2013). For example, the species of the genus *Difflugia* are xenosomic (agglutinating particles from the environment to build their test) and live in wet conditions of oligotrophic peatlands, likely because the silica particles they use for their tests are more abundant in shallow pools (Meisterfeld 1977, Mitchell et al. 2008, Schonborn 1962). Other observational studies have shown that a correlation may exist between seasonal changes in soil moisture content and Arcellinida abundance and community composition (Heal 1964, Mitchell et al. 2008, Warner et al. 2007). Likewise, relationships with pH and macro-nutrients have been found (Tolonen et al. 1992, Mitchell et al. 2008). It has further been suggested that species size and shape may dictate in which environments they will best survive. For example, thicker water films on plant surfaces in higher soil moisture may permit many species to reproduce, while thinner water films in drier soils may better suit larger amoebae (Mitchell et al. 2007, McKeown et al. 2019).

The shape and composition of the tests are highly diverse and test size ranges from 20-300 µm across the ~1100 described species of Arcellinida (Meisterfeld 2002, Steele et al. 2018). In some species, tests are either made from calcareous, proteinaceous, or siliceous material, whereas other species “glue” together organic and/or mineral particles from the environment to build agglutinated tests (e.g. Meisterfeld 2002, Mitchell et al. 2008, Ogden and Hedley 1980). Given the richness of test characteristics, Arcellinida species have been described using morphological features that can be observed by light microscopy (e.g. Ogden and Hedley 1980, Meisterfeld 2002). A comprehensive morpho-taxonomy has been established, which serves as the basis for most bioindicator and fossil record studies and relies on differences in test size,

composition and shape for species delineation. Genetic studies of the last decade on the one hand allowed the disentanglement of closely related species of *Nebela* (Lara et al. 2008, Kosakyan et al. 2012, Heger et al. 2010, Charman 1999, Kosakyan et al. 2013, Singer et al. 2015), yet, on the other hand revealed the presence of cryptic diversity within the defined morphospecies, suggesting that Arcellinida diversity has been underestimated (Kosakyan et al. 2012). Some cryptic species have been observed to have different ecological requirements (e.g. Singer et al. 2015), thus putting into question the usefulness of the morphospecies concept for this group of microbial eukaryotes. For example, genetic analyses of variable regions in the SSU-rDNA as well as in the mitochondrial cytochrome oxidase gene subunit 1 (COI) have resulted in the discovery of up to 12 genetic types within the morphospecies *Hyalosphenia papilio*, which show more restricted distribution patterns compared to the morphospecies (Singer et al. 2019, Heger et al. 2013, Oliverio et al. 2014).

The existence of cryptic diversity and phenotypic plasticity limits the effectiveness of light microscopy for understanding the diversity and activity of these organisms (Oliverio et al. 2014, Kosakyan et al. 2012). Therefore, for this pilot study, we designed Arcellinida-specific primers for high throughput sequencing (HTS) analyses, and deployed them to explore the diversity of Arcellinida in both vertical cores and horizontal transects in a New England fen ("Hawley Bog", Hawley, MA, USA). Hawley Bog is managed jointly by the Nature Conservancy and Five College Inc (https://www.fivecolleges.edu/community/hawley-bog), and we sample on the floating mat where we typically find pH in the range of 4-5. We designed primers for the V6 and V7 variable regions of the SSU-rDNA gene by identifying areas conserved in Arcellinida but distinct from other Amoebozoa. We then tested these primers on DNA (active and quiescent community) and RNA (active community and indication of biomass) extracted from *Sphagnum* moss samples, amplified the fragment by PCR and sequenced it by Illumina MiSeq sequencing. We analyzed the resulting amplicons using a custom-built bioinformatic pipeline.

MATERIALS AND METHODS

Sample collection and preparation

We sampled at two different spatial scales: 36 cm transects along the surface and in a core (Experiment 1) to provide a snapshot of the local Arcellinida community and 26 m surface transects (Experiment 2) to provide insight into the Arcellinida community across a broader landscape. For Experiment 1, we collected a total of four 36 cm moss segments at a local fen in Hawley, MA (42.575534 N, 72.890580 W): two horizontal segments along the surface of the fen (18 August 2017) and two vertical cores using a Russian box corer ((50 cm x 10cm x 10cm); 27 July 2017; 18 August 2017). One of the 36 cm surface transects abutted the second depth transect, while the other was 1 m away from the second depth transect. We divided each segment into six equally sized slices, and washed amoebae from moss using the protocols described below. Due to the challenges associated with pushing the box corer through the dense surface mat of *Sphagnum*, we sampled the vertical cores in areas with less vegetative cover (i.e. less robust *Sphagnum* matt), and so we treat the surface of the core separately from other surface samples.

To conduct morphological observations in Experiment 1, we prepared Petri dishes from the 24 slices of the horizontal segments and the vertical cores by washing amoebae from ~10 strands of *Sphagnum* (surface) or 10 grams of peat (core) in 50 mL conical tubes using ~30 mL filtered (0.2 µm filter) *in situ* water. The tubes were inverted ~10 times to wash the amoebae off the moss. Within a few days of collecting, we counted the numbers of tests (both living and empty) of the genera *Arcella, Heleopera, Hyalosphenia,* and *Nebela* s.l. identified in each plate over a 20-minute period using light microscopy (inverted microscope Olympus CKX41 with 200x magnification). We focused on the taxonomic rank of genus in order to ensure a high consistency among researchers.

To isolate cells for molecular analyses for Experiment 1, we washed ~10 g of moss by first filtering over a 300 µm filter to remove plant matter and then over a 35 µm filter to wash away bacteria (as well as smaller amoebae). The filtrate smaller than 300 µm and larger than 35 µm was spun at ~580g for one minute to pellet any cell material and stored in 1 mL RLT buffer (Qiagen, Germantown, MD, USA) at -80ºC until DNA or RNA extraction (see below).

To refine our insights into Arcellinida distribution within the fen, as well as to gain an understanding of the size distribution of OTUs and potentially of different life stages, we conducted Experiment 2 by sampling on the 6 June 2018. For Experiment 2, we sampled a handful of moss at each of 7 stations along two 26 m surface transects, and then focused on every other station to capture those spanning from the edge of the open fen at the forest (Station 1) through the open fen (Stations 3 and 5) to the edge along the open water (Station 7). These stations were 8 m apart from each other.

For each sample, we washed 50 g of moss in 250 mL of filtered (0.2 µm filter) *in situ* water, and filtered the sample by gravity through three consecutive meshes: a 300 µm mesh, a 100 µm mesh, a 35 µm mesh, and then using a vacuum pump for the 10 µm and the 2 µm polycarbonate membranes. For the larger sizes (100µm and 35 µm), each mesh was rinsed with filtered *in situ* water to allow smaller cells to pass through, but we did not increase the pressure to avoid cell breakage during filtration. For the two smaller sizes, we filtered until the filter clogged and used a maximum of 2 filters per size and sample. Each filter (100 µm, 35 µm, 10 µm, 2 µm) was directly stored in 1 mL of RLT buffer (Qiagen, USA) at -80ºC until DNA or RNA extraction (see below), and only the 300 µm mesh was reused after washing between experiments. Hence, we evaluated the filtrates in these four ranges: 300-100 µm, 100-35 µm, 35-10 µm, and 10-2 µm and discarded any material that would not pass through the 300 µm filter or that passed through the 2 µm membrane.

Community DNA and RNA Extractions

For the extraction of whole community DNA (i.e. active plus quiescent), we used the ZR Soil Microbe DNA MiniPrep™ extraction kit (Zymo Research, Tustin, CA, USA) and for RNA (i.e. active community members and indication of biomass) we used the Qiagen RNEasy Mini kit (Qiagen, USA) following Sisson et al. (2018). In both cases, we followed the manufacturers' protocols. Before RNA extraction, the RNA hood and pipettes were treated with RNase away (Thermofisher, Waltham, MA, USA). We further processed the RNA extractions with the TURBO DNA-free™ Kit (Thermofisher, USA) to remove DNA from the extracted RNA and, afterwards the SuperScript® III First-Strand Synthesis System (Thermofisher, USA) with random hexamer primers (Thermofisher, USA) to generate single-stranded cDNA. The DNA and cDNA were stored at -80ºC until further processing.

Primer Design

We designed PCR primers for two loci – SSU-rDNA and cytochrome oxidase 1 (COI) – to specifically amplify Arcellinida while avoiding other Amoebozoa (and other eukaryotes). To this end, we retrieved Amoebozoa SSU-rDNA and COI sequences available on GenBank on the 14 August 2017 (Table S1). We then used MAFFT (Katoh and Standley 2013) to initially align the named sequences and FastTree (Price et al. 2009, Price et al. 2010) to build a tree. EMBOSS Water (Rice et al. 2000) allowed the removal of identical sequences by creating pairwise alignments of sequences close on the tree. We generated 80% similarity consensus sequences from designated in-group and out-group sequences using Mesquite (Maddison and Maddison 2011). The in-group was composed of Arcellinida sequences while the outgroup included other Amoebozoa (i.e. Mycetozoa and Discosea). For all primers, the last nine bases of each primer exactly matched the consensus sequence while we resolved ambiguities in the 5' end of the primer to balance GC content and avoid hairpins (Table S2). We ultimately chose primers designed to match a region of the V6 and V7 hypervariable regions of the SSU-rDNA, conserved in the in-group, for the focus of the work (primer set 2088+/2435-; Table S3).

Primer assessment

For an initial primer test, we deployed the primers on single-cell whole genome amplifications (WGAs) of the Arcellinida species *Hyalosphenia elegans*, *Hyalosphenia papilio* and *Arcella hemisphaerica*. We ran gradient PCRs to identify the primers' ideal annealing temperatures (Tables S3). For polymerase chain reaction (PCR), we used Q5® Hot Start High-Fidelity DNA Polymerase (NEB, Ipswich, MA, USA). For each reaction, we used 19 µL of master mix (14 µL H_20, 4 µL Q5 5x buffer (NEB, USA), 0.4 µL 100 mM dNTPs, 0.2 µL of each primer (200 pM), and 0.2 µL Q5 Hot Start High Fidelity polymerase (NEB, USA)) and 1 µL sample DNA. PCR products were cleaned using Agencourt AMPure XP Beads (Beckman Coulter, Brea, CA, USA) and prepared for direct sequencing using the BigDye Terminator v3.1 Cycle Sequencing Kit (Thermofisher, USA) and then cleaned using the Performa DTR Gel Filtration Cartridges (Edge BIO, Gaithersburg, MD, USA). Sanger sequencing was performed at Smith College Center for Molecular Biology. The sequencing results were assessed with SeqMan v.12.0 (DNASTAR) and compared to previously recorded WGA sequencing results that had been generated using Arcellinida-specific PCR primers from Oliverio et al. (2014).

To further confirm the primers' ability to bind to Arcellinida sequences, we deployed them on a few community samples to generate amplicons in several pilot experiments. The SSU-rDNA primer sets worked robustly on multiple samples, while the COI primer results were more variable. We cloned the resulting SSU-rDNA PCR products using the Zero Blunt Topo PCR Cloning Kit (Thermofisher, USA) with One Shot TOP10 chemically competent *E. coli* cells (Thermofisher, USA). Using Hot Start Taq DNA polymerase, Agencourt AMPure XP Beads, BigDye Terminator v3.1 Cycle Sequencing Kit, and Performa DTR Gel Filtration Cartridges, we sequenced cloned amplicons from colonies to assess their ability to capture diverse Arcellinida. These pilot analyses suggest that the SSU-rDNA primers 2088+/2435- successfully

characterize Arcellinida taxa; however, they also annealed to closely related amoebozoan lineages and select ciliates and fungi. We report our determination of best temperatures for all primers in Table S3.

High-Throughput Amplicon Sequencing

Based on our determination of primer effectiveness from pilot studies described above, we focused on SSU-rDNA amplicons generated by the 2088+/2435- primer pair for the analysis of the DNA and RNA samples from Experiments 1 and 2. The PCR cycling conditions began with a denaturation step at 98°C for 3 min, followed by 37 cycles of 98°C for 15 s, 64°C for 30 s, 72°C for 1:30 min, and a final extension at 72°C for 10 min. We diluted the initial PCR products 1:50 and 1:10 according to their band brightness and performed a second five cycle PCR using the dilutions to add the Illumina adapters (Forward primer 2088 with adapter: 5' – TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGAAACTCACCAGGTCMRGACAC – 3'; Reverse primer 2435 with adapter: 5' – GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCATCACAGACCTGTYKTKGCC – 3). We conducted each PCR in triplicate and pooled the products to reduce PCR bias (Jung et al. 2012, Lahr and Katz 2009). We used Agencourt AMPure XP Beads to remove primer dimer from the combined amplicons. The University of Rhode Island Genomics and Sequencing Center conducted Illumina MiSeq 2X150 High-Throughput Sequencing on the generated amplicons.

Data analyses

We relied on methods described in Sisson et al. (2018) to trim reads, remove chimeras and call OTUs using a SWARM distance of 1 (Mahe et al. 2014, Mahe et al. 2015). We then used a three-step approach to identify and remove outgroup OTUs: 1) a comparative BLAST; 2) a

phylogenetic analysis using non-Amoebozoa outgroups and 3) finally, a phylogenetic analysis using just Amoebozoa taxa. First, we used BLAST (Altschul et al. 1990) to compare each OTU to both a database of Amoebozoa sequences (Supplemental file S1) and a database of outgroup sequences from GenBank (Supplemental file S2). We then removed sequences with an E-value ratio (E-value of the best Amoebozoa BLAST divided by the E-value of the best outgroup BLAST) higher than 10^{15}. We used this criterion to compensate for the paucity of the Arcellinida SSU-rDNA database as otherwise our preliminary analyses suggested we would retain long-branch outgroup taxa. In the second step, we combined the retained sequences with previously-identified Arcellinida plus outgroups from Amoebozoa plus all major eukaryotic clades from GenBank, which we provide as a fasta file and constrained tree (Supplemental files S3 and S4, respectively). We aligned the sequences using the L-INS-i option in MAFFT (Katoh and Standley 2013) to account for variation in length, and then placed our OTUs into a constrained tree using the Evolutionary Placement Algorithm (Berger et al. 2011) implemented in RAxML v8.0 (Stamatakis 2014). This algorithm allows for the addition of short sequences (usually the case in HTS) to create a robust tree built with full length SSU-rDNA. In the final step, OTUs that fell within Amoebozoa were added to a second file that contained only Amoebozoa, and this file is provided as a fasta file (Supplemental file S5). We again aligned with L-INS-i and constructed a phylogeny with RaxML to identify the ingroup taxa, and we show this topology in the results section. The reads from ingroup taxa were randomly subsampled (or rarefied) at 35,000 reads per sample to avoid bias in diversity assessment related to depth of sequencing.

The impact of the geographical distribution, size and activity (DNA vs RNA) on the community composition were assessed using the UniFrac dissimilarity index (Lozupone and Knight 2005, Lozupone et al. 2011) and Principal Coordinates Analysis (PCoA) using the Phyloseq (McMurdie and Holmes 2013) and vegan packages (Oksanen et al. 2016) implemented in R.

RESULTS

Primer design

Though we worked to develop primers for two markers, SSU-rDNA and cytochrome oxidase 1, we focus our analyses on the set of SSU-rDNA primers that worked consistently across varying Arcellinida whole genome amplifications and communities (i.e. 2088+/2435-; Table S3). We provide all of our primers with some details on their effectives (Tables S1-S3), as all worked for at least some taxa, and we invite potential users to do their own bioinformatic comparisons before trying on their target taxa. This is particularly critical as the paucity of data at the time of our primer design will surely be ameliorated by the community through genetic analyses of morphospecies of Arcellinida.

Arcellinida diversity estimated with SSU-rDNA primers

We conducted two studies using our focal primer set 2088+/2435-: the first focusing on small scale cores (9 samples) and surface transects (12 samples), and the second only on surface samples. Experiment 1 is a small-scale study of two 36 cm cores and two 36 cm surface transects that focused on active Arcellinida (i.e. RNA) for which we also collected morphological data on a few genera. Experiment 2 included a larger-scale transect (i.e. 26 m meters along the surface) for which we first size-fractioned the community (i.e. 100-300 μm, 35-100 μm, 10-35 μm, 2-10 μm; eight DNA and eight RNA samples each, except 10-35 μm fraction, which has only six DNA samples) and then analyzed both total (i.e. DNA; 30 samples) and active (i.e. RNA; 32 samples) communities. The bulk of the community DNA and RNA amplified well with our primers, yet a few samples did not amplify reliably and so are not included in the analyses (e.g. some of the RNA samples from core 1 of Experiment 1 and some DNA samples from Experiment 2).

Across both experiments, the 2088+/2435- primer set captured a total of 317 OTUs that fall among the Arcellinida rDNA sequences available on GenBank (Table S4, with best BLAST hit listed). As described in our methods, these 317 OTUs fall among Arcellinida after passing three filters: comparative BLAST analysis, a phylogenetic analysis including non-Amoebozoan outgroups (Supplemental file S6, with 272 OTUs that were removed), and then a phylogenetic analysis of only Amoebozoa (Supplemental file S7, with 137 OTUs that were removed). This large number of non-target OTUs is not surprising given the patterns of conservation in the ssu-rDNA locus, and might be reduced by further experiments to optimize PCR conditions (i.e. annealing temperature, cycle number).

Among the 317 OTUs retained, three OTUs dominated the ~3 million rarefied reads in our studies: OTU1 (848,241 reads, 29.6%), OTU3 (792,462 reads, 27.6%), and OTU2 (674,768 reads, 23.5%; Table S4). Only four other OTUs represented more than 1% of total reads: OTU4 (272,854 reads, 9.5%), OTU16 (48,860 reads, 1.7%), OTU18 (34,980 reads, 1.2%), and OTU31 (32,386 reads, 1.1%, Table S4). In recognition that our rare OTUs may be the result of experimental error and/or somatic mutations within amoebae, we focus the remaining analyses on the 92 OTUs that have a minimum of 100 reads and that occur in at least nine samples (Table 1). For each of these 92 OTUs, we assessed the relative read numbers (e.g. >100, >1,000, >10,000), whether they are enriched (>80% reads) in core versus surface in Experiment 1 and how they are distributed across sizes in Experiment 2 (Fig. 1).

Using position on our phylogeny, we either assigned the 92 OTUs (with more than 100 reads and in at least nine samples) to previously-characterized species or genera, or designated them as Arcellinida spp. (Fig. 1 'Novel Arcellinida clade', Table 1). Perhaps the most striking pattern in the phylogenetic analyses of these 92 OTUs is the presence of a clade of 21 'novel' OTUs that are distinct from any previously DNA barcoded lineages but represent only 0.4 % of our

total reads (i.e. ~12K out of ~2.8 million reads; Table 1 and Fig. 1). We also found 18 OTUs (2.1% of total reads) that cluster close to *Nebela penardiana* (now: *Longinebella penardiana*), and 10 that cluster within clades containing *Hyalosphenia papilio* or *Hyalosphenia elegans* (24.1% and 28.1% of reads, respectively; Table 1). Eight OTUs designated as Arcellinida fall in various places outside of characterized clades, and the remaining 27 OTUs match nine characterized taxa (Table 1, Table S5).

Based on the sum of read numbers, the OTUs closely related to four morphospecies dominate our samples, accounting for 2,617,659 of 2,839,139 reads (92.2%): *Physochila griseola* (859,064 reads), *Hyalosphenia papilio* (796,627 reads), *Hyalosphenia elegans* (684,684 reads), and *Arcella hemisphaerica* (277,384 reads; Table 1). All remaining taxa represent less than 2.5% of total reads (Table 1).

Analyses of patterns among 92 focal OTUs

We assessed the presence of taxa between cores versus surface (Experiment 1, 36cm transects; 9 vs. 12 samples, respectively) and between DNA and RNA (Experiment 2, 26m transects; 30 vs. 32 samples, respectively), and highlight those taxa for which more than two thirds of reads were found in one treatment (Table 1, column 3). Here, we disregarded the one OTU that matched *Argynia dentistoma* as its low read number (430 reads total) prevented us from drawing conclusions. In Experiment 1, seven clades were enriched in the surface fraction (Novel Arcellinida Clade (0.70), *Hyalosphenia papilio* (0.72), *Arcella hemisphaerica* (0.89), *Nebela* spp. (0.75), *Bullinularia* sp. (0.96), *Centropyxis* sp. (0.73), and *Difflugia bacillariarum* (0.86; Table 1). In the case of *D. bacillariarum,* where our OTU is identical to a published sequence (GB# JQ366065) the location on the surface is consistent with its close association with diatoms as evidenced in a previous study of bogs in Europe (Gomaa et al. 2012). Three clades are enriched in the core: *Physochila griseola* (0.75), *Difflugia* spp. (0.91), and

Cornutheca ansata (0.68) while three were more evenly divided between core and surface: *Nebela penardiana* (now: *Longinebella penardiana*)*, Hyalosphenia elegans,* and the non-monophyletic Arcellinida spp. (Table 1).

We also found three taxa enriched in either the total community (i.e. amplified from DNA) or active community (i.e. amplified from RNA) in the surface-only samples of Experiment 2. The three lineages with more than two thirds of their reads from DNA are the novel Arcellinida clade (0.81), *Physochila griseola* (0.80), and *Difflugia* spp. (0.76; Table 1). In contrast, three morphospecies appear to be relatively highly active: *Hyalosphenia elegans* (0.70) and *Hyalosphenia papilio* (0.74) and *Arcella hemisphaerica* (0.81; Table 1). Reads for the remaining taxa are roughly evenly distributed between total (DNA) and active (RNA) proportions (Table 1).

Comparison to morphological analysis of Arcellinida diversity

Though most of the lineages observed by microscopy are also retrieved by HTS, a quantitative comparison of the two is not possible given the inherent bias in molecular data and the investment of effort required to ID morphospecies (i.e. training and observing). Nevertheless, we did assess broad-scale patterns for the few lineages that we could identify with confidence: the genera *Heleopera, Hyalosphenia, Nebela* s.l. and *Arcella*. Across the Experiment 1 surface transects and a portion of the core transects, the proportion of *Hyalosphenia* tests identified are similar to the proportion of *Hyalosphenia* sequences identified via HTS (Table S6). More variability existed between the abundance of *Arcella* and *Nebela* tests observed under the microscope and the number identified via HTS. While we observed *Heleopera* sp. to be present in the samples under the microscope, we did not detect this genus in the molecular analysis.

Analyses of the community based on activity levels, depth and size

We assessed presence/absence patterns for the entire 317 OTU dataset (Table S4) using principle coordinates analyses (PCoA) with UniFrac distances to account for evolutionary relatedness. Comparing all samples from Experiments 1 and 2, we see an intriguing pattern as communities generally divide into one of two groups: group 1 = surface RNAs (Experiments 1 and 2) and group 2 = surface DNA (Experiment 2) and core RNA (Experiment 1; Figs. 2 and S1). In other words, the active communities below the surface are similar to the total communities on the surface, and distinct from the active community on the surface (Fig. 2 and S1).

We also evaluated the difference among communities looking at the active members (RNA) of surface vs. core samples for Experiment 1 (Fig. 3). Here, PCoA analyses explained over 50% of variation (30.2% and 20.9% on the x and y axis, respectively; Fig. 3) and showed a clear pattern of clustering by sample type: all surface samples cluster together, while the two core samples form distinct clusters (Fig. 3). Depth within sample is not a clear driver of the pattern as the community composition for core 2 is distinct from core 1 (Fig. 3).

For Experiment 2, we used the same PCoA approach to evaluate the pattern of active (RNA) communities by size and found that the bulk of the variation (26.7%, x axis Fig. 4) distributes from the smallest to the largest size classes. Along the y-axis (11.5% of variation), the greatest distinction appears between the two largest size fractions (100-300 μm (red) and 35-100 μm (orange)) from site 1 of both transects 1 and 2. Intriguingly, this site is at the forest/fen boundary.

We further looked at size discrimination among our 92 focal OTUs, as opposed to the entire community analyzed in the PCoAs. A total of 23 OTUs had over half of their reads fall within one particular size class (e.g. OTU4, OTU280, OTU169, OTU476, OTU417, OTU973, OUT705,

OTU255; Fig. 1, Table S5). For instance, 88 percent of the reads for OTU705 fell within the 2-10 µm size class. Likewise, 85 percent of OTU169 and OTU417 are within the 10-35 µm and the 100-300 µm size class, respectively. These data, combined with the fact that only the three most abundant OTUs are present in all size classes (likely due to their abundance plus presence in aggregates), are evidence of the efficiency of our fractionation method.

Many OTUs are found in two or three size classes (e.g. OTU2, OTU16, OTU338; Fig. 1, Table S5). An additional seven OTUs occur within two disjunct size classes, most often the 2-10 µm and the 100-300 µm size classes (i.e. OTU29, OTU356, OTU575, OTU176, OTU365, OTU570, OTU607; Fig. 1, Table S5*)*. In two instances, disjunct OTUs fall sister to one another on the tree (i.e. OTU356 and OTU575, closely related to *Physochila griseola*) and there is a clade of three OTUs within *Arcella* that are all disjunct, with OTU365 and OTU176 found in both 2-10 µm and 100-300 µm fractions and OTU405 in 10-35 µm and 100-300 µm.

Spatial patterning:

Even based on the limited sampling here, we have evidence that this amplicon-based approach will be useful for discerning small-scale patterns as a subset of OTUs are enriched for specific locations within the fen, defined here as having >90% of reads in a single area. For example, in our active communities (i.e. RNA) in Experiment 2 and among the 92 focal OTUs, four are found nearly exclusively at station 1 (S1 in both transect 1 and 2), which falls just at the edge of the fen under small trees: OTU847 (within novel Arcellinida clade, 100% of reads), OTU440 (*H. elegans,* 98% of reads), OTU45 (*Cornutheca ansata,* 96% of reads) and OTU883 (novel Arcellinida clade, 92% of reads; Fig. 1, Table S5). Four OTUs are enriched within the open fen (i.e. not forest or water edges; stations S3 and S5): OTU510 (Arcellinida sp., 100% reads), OTU530 (novel clade, 100% reads), OTU903 (*H. papilio* clade, 100% reads), OTU647 (*H.*

papilio, 95% reads; Fig. 1, Table S5). Six lineages are enriched at the edge of the fen that abuts an open stream (i.e. stations 7): OTU582 (*P. griseola*, 100% of reads), OTU1325 (*N. penaridana*, 100% reads), OTU787 (Difflugia sp.) and three OTUs within the novel Arcellinida clade (OTU904 (100% reads), OTU503 (94% reads) and OTU344 (92% reads; Fig. 1, Table S5). Reads for many other OTUs were found spread across the fen in a patchy manner (Table S5).

DISCUSSION

Choice of locus and region

We chose to focus on primers 2088+/2435- for the SSU-rDNA locus as its mixed patterns of conserved and variable regions provided the best opportunity for developing primers for all Arcellinida, a clade that is at least 700 million years old (Porter and Knoll 2000). We acknowledge that work in our lab and others has demonstrated the greater power of cytochrome oxidase I (COI) as a phylogenetic marker for this group to resolve diversity on the species level (e.g. Oliverio et al. 2014, Kosakyan et al. 2012). However, the lack of conservation in nucleotide sequences (i.e. the highly variable 3rd positions in this protein coding gene) plus the relatively limited sampling of COI sequences on GenBank made primer design very difficult and our best efforts (Tables S2 and S3) worked only on some but not all taxa. We provide the COI sequences for any interested in testing them out, and we also acknowledge that the multiple copies of the rDNA locus within genomes may confound inferences about closely related lineages. However, our goal here is not to reconstruct phylogeny but instead to compare Arcellinida across habitats, and the variable regions we chose (V6 and V7) are sufficient for this purpose (see Fig. 1). As for regions of the SSU-rDNA, we are aware of the 'popularity' of some regions (e.g. V4 and V9), but believe discussions on what is the 'best' region often lose track of the goal, which is to find the most variable region that works for our target taxa. We chose to

focus on V6 and V7 because this was the best place to design primers that allowed robust classification of resulting sequences.

Inferences about Arcellinida diversity

Across both experiments, we identified a total of 317 OTUs that fall within Arcellinida after filtering for quality and removing outgroup sequences (e.g. from *Echinamoeba, Flabelluidae. Nolandella, Leptomyxa,* see yellow highlights in Table S4). Of these 317 OTUs, 92 OTUs were of high enough abundance (>100 reads) and occurrences (>9 samples) to warrant further investigation (Table 1, Table S2, Fig. 1). We found the greatest diversity of lineages (i.e. 21 OTUs) in a clade that lacked identified representatives on GenBank ("Novel Arcellinida" clade in Fig. 1). We suspect that these lineages represent morphologically well-characterized genera (e.g. Meisterfeld 2002) that await molecular analyses. They could also represent small inconspicuous, to date undescribed species, as they have been observed to exist in Euglyphida, another group of testate amoebae within the Rhizaria (Tarnawski and Lara 2015). However, this may be less likely as we found some of these lineages also in the larger size fractions in our filtration experiment (e.g. OTU512, Fig. 1, last column).

Compared to the limited morphological data we obtained in the 36 cm core/surface transects in Experiment 1, we saw some correspondence to our estimates of the genus *Hyalosphenia*, but considerable variation with counts of *Arcella* and *Nebela* s.l. (Table S6). Most likely this discrepancy is due to the fact that we counted tests (empty or not), and only for a short time per sample (i.e. 20 minutes) while our molecular data estimate total living (active and encysted) Arcellinida. Other contributing factors to discrepancies between molecules and morphology include variation in rDNA copy number, which is common among eukaryotes (Wang et al. 2019, Wang et al. 2017, Huang and Katz 2014) and/or affinity of primers. To our surprise, we found

no molecular evidence of *Heleopera*, though only two sequences from this genus were available on GenBank at the time of writing. It is possible that nucleic acids were not isolated from this species if the cells did not lyse in buffer, or that our primers may not be well matched for this genus, and/or that Group I introns or other inserts may generate length variation that is selected against in our community PCRs (Lahr et al. 2011, Gomaa et al. 2012).

Comparing patterns among community samples

We detected noticeable differences between active (RNA) and total (DNA) communities, with surface active communities tending to cluster distinct from core RNAs and surface DNAs. DNA has been the gold standard to characterize environmental microbial community diversity in the last two decades (e.g. Schmidt et al. 1991, Amann et al. 1995). Compared to DNA, RNA is less stable, and analyzing RNA requires an extra step to generate cDNA (reverse transcription) for the downstream analysis. Nevertheless, RNA provides a proxy for the active community, at least at the specific time of isolation (Blazewicz et al. 2013). Both DNA and RNA isolation have limits in analyses of communities including differential copy number within organisms (Klappenbach et al. 2001, Molin and Givskov 1999), the presence of RNA in dormant cells and cysts (e.g. Mahe et al. 2017, Worden and Binder 2003), which suggest that rRNA is not directly synonymous of activity but more of 'protein synthesis potential' (Blazewicz et al. 2013). Nevertheless, it has been shown for phytoplankton that RNA is a better predictor of biomass or carbon content than DNA (Maki et al. 2017). In our study, we recognize that the observed differences between communities assessed by DNA and by RNA may be related to biomass (higher proportion in RNA suggesting higher biomass) and/or to protein synthesis potential (Fig. 2).

Even in our limited study, we found spatial patterns as core communities are distinct from surface (Experiment 1; Fig. 3) and forest communities are distinct from open fen (Experiment 2;

Fig. 4). At the level of OTUs (a proxy for species or populations) we also found lineages with distinct preferences for forest, open fen, or water's edge in our 26 m transects (Experiment 2, see results). There has been more work on testate amoebae in bogs and fens in Europe than in North America (e.g. Lamentowicz et al. 2013, Jassey et al. 2013, Jassey et al. 2011, Mitchell et al. 2008, Lamentowicz and Mitchell 2005). The testate amoeba community composition has been shown to be variable both on a horizontal and vertical perspective in bogs and fens (e.g. Roe et al. 2017, Mitchell et al. 2000). The presence of certain testate amoebae morphospecies does vary with ecological factors (e.g. pH, depth to water table, season, etc.), making them useful as bioindicators in paleoecological and contemporary studies (Booth et al. 2008, Markel et al. 2010, Jassey et al. 2013, Lamentowicz et al. 2013, Payne 2013, Song et al. 2018). Intriguingly, the surface DNA and core RNA are not markedly different based on a PCoA (Figs 2 and S1). In fact, two of the three taxa with a higher contribution in those samples are shared (*Physochila griseola* and *Difflugia* spp.; Table 1). Using morphology, some species within the genus *Difflugia* are found associated with sediment and root (Lahr and Lopes 2006), though the limited molecular data from this clades makes direct comparisons to our study difficult. Similarly, *Physochila griseola* was found to dominate in the deeper part of a six-centimeter vertical profile (Song et al. 2016). This is in accordance with our observation of higher proportion in the RNA core samples. The discovery of diverse Arcellinida within the peat core is surprising and may hint towards the existence of lineages that can withstand low oxygen. Alternatively, we may have captured preserved environmental nuclei acids, though our filtration system may have removed the bulk of this material.

Size variation within and between OTUs suggest complex life histories

Our approach to filtering by size revealed intriguing patterns as communities separated from largest (100-300 µm) to smallest (2-10 µm) size classes (Fig. 4) and, perhaps more intriguingly, some OTUs appeared distinct (i.e. present in 2-10 µm and 100-300 µm, but not in the

intervening size; Fig. 1). Only our most abundant three OTUs appeared in all of the size classes, potentially due to morphological plasticity but more likely just due to their high abundance such that they appeared as debris in all sizes (Fig. 1). Most other OTUs have continuous ranges that are consistent with the biology of their closest morphospecies: OTUs in the *Nebela penardiana* (now: *Longinebela penardiana*) clade range from 2-100 µm, consistent with the published size of around 80 µm in *N. penardiana* var minor and slightly larger than 100 µm for *N. penardiana* var penardiana (Lara et al. 2008). An advantage of our approach is that likely the bulk of environmental DNA was washed out as our smallest pore size was 2 µm.

Several explanations exist for the disjunct size distributions of some OTUs (i.e. OTU29, OTU176, OTU356, OTU365, OTU405, OTU570, OTU575, OTU607; Fig. 1). The regions amplified by these primers may fail to resolve the identity of these lineages, they may be smaller species that get trapped in debris and/or in larger organisms (prey-predation relationship or symbiosis), or these data could potentially be capturing different life stages of the specific testate amoebae (e.g. adults vs. gametes). Data on the life cycles of Arcellinida are sparse, with only a single report of meiotic structures during cyst formation in *Arcella vulgaris* (Mignot and Raikov 1992) and a single report of small 'naked' life history stages in the same species (Cavallini 1926). Volkova and Smirnov (2016) showed that *A. vulgaris* can survive when extracted from test, and some naked cells did generate new tests, but data on other genera are lacking.

Synthesis

Here, we report on our development of Arcellinida-specific SSU-rDNA primers for community analyses on the molecular level and provide insights from pilot sampling from multiple transects in a fen (Hawley Bog, Hawley MA). Our analyses demonstrate the potential of this approach for characterizing active (i.e. RNA-based analyses) and total (i.e. DNA-based analyses) Arcellinida community members, and for identifying community patterns based on location (i.e. Experiment 1: 36 cm transects along the surface and within cores) and size (i.e. Experiment 2: 26 m surface transects fractioned to 100-300 µm, 35-100 µm, 10-35 µm, 2-10 µm). We acknowledge that our primers, like all HTS primers, are biased by both the availability of taxa at the time of design and the kinetics of PCR, and we see the strengths in deploying these primers with our bioinformatic pipeline in comparing among samples rather than in capturing all Arcellinida. Taken together, the tools developed here will be of use to others studying Arcellinida communities across diverse habitats, including studies of Arcellinida as bioindicators in changing environments.

ACKNOWLEDGEMENTS:
We are very grateful to Gabrielle Östlund-Sholars, Naomi Ostriker, Evie Hoffman, and Olivia Dufour for their assistance with sampling and primer testing. We extend our appreciation to Janet Atoyan at the University of Rhode Island for running the MiSeq sequencing of our samples. We are also very grateful for the detailed comments of two anonymous reviewers.

Funding: This work is supported by an NSF grant DEB-1651908 to LAK.

Data archiving: The raw reads are available from GenBank under the BioProject PRJNA563881.

LITERATURE CITED

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *Journal of molecular biology,* **215:**403-410.

Amann, R. I., Ludwig, W. & Schleifer, K. H. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev,* **59:**143-69.

Berger, S. A., Krompass, D. & Stamatakis, A. 2011. Performance, Accuracy, and Web Server for Evolutionary Placement of Short Sequence Reads under Maximum Likelihood. *Systematic Biology,* **60:**291-302.

Blazewicz, S. J., Barnard, R. L., Daly, R. A. & Firestone, M. K. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *The ISME journal,* **7:**2061.

Bobrov, A. A., Andreev, A. A., Schirrmeister, L. & Siegert, C. 2004. Testate amoebae (Protozoa: Testacealobosea and Testaceafilosea) as bioindicators in the Late Quaternary deposits of the Bykovsky Peninsula, Laptev Sea, Russia. *Palaeogeography Palaeoclimatology Palaeoecology,* **209:**165-181.

Booth, R. K. 2001. Ecology of testate amoebae (Protozoa) in two lake superior coastal wetlands: Implications for paleoecology and environmental monitoring. *Wetlands,* **21:**564-576.

Booth, R. K., Sullivan, M. E. & Sousa, V. A. 2008. Ecology of testate amoebae in a North Carolina pocosin and their potential use as environmental and paleoenvironmental indicators. *Ecoscience,* **15:**277-289.

Bosak, T., Lahr, D. J. G., Pruss, S. B., Macdonald, F. A., Dalton, L. & Matys, E. 2011. Agglutinated tests in post-Sturtian cap carbonates of Namibia and Mongolia. *Earth and Planetary Science Letters,* **308:**29-40.

Cavallini, F. 1926. The asexual cycle of development in *Arcella vulgaris*. *Journal of Experimental Zoology,* **43:**245-255.

Charman, D. J. 1999. Testate amoebae and the fossil record: issues in biodiversity. *Journal of Biogeography,* **26:**89-96.

Creevy, A. L., Andersen, R., Rowson, J. G. & Payne, R. J. 2018. Testate amoebae as functionally significant bioindicators in forest-to-bog restoration. *Ecological Indicators,* **84:**274-282.

Finlay, B. J. 2002. Global dispersal of free-living microbial eukaryote species. *Environmental Microbiology,* **296:**1061-1064.

Finlay, B. J. & Clarke, K. J. 1999. Ubiquitous dispersal of microbial species. *Nature,* **400:**828-828.

Finlay, B. J., Esteban, G. F., Clarke, K. J. & Olmo, J. L. 2001. Biodiversity of terrestrial protozoa appears homogeneous across local and global spatial scales. *Protist,* **152:**355-366.

Finlay, B. J., Esteban, G. F., Olmo, J. L. & Tyler, P. A. 1999. Global distribution of free-living microbial species. *Ecography,* **22:**138-144.

Foissner, W. 1999. Protist diversity: Estimates of the near-imponderable. *Protist,* **150:**363-368.

Gilbert, D., Amblard, C., Bourdier, G. & Francez, A. J. 1998. The microbial loop at the surface of a peatland: Structure, function, and impact of nutrient input. *Microbial Ecology,* **35:**83-93.

Gomaa, F., Todorov, M., Heger, T. J., Mitchell, E. A. D. & Lara, E. 2012. SSU rRNA phylogeny of Arcellinida (Amoebozoa) reveals that the largest arcellinid genus, Difflugia Leclerc 1815, is not monophyletic. *Protist,* **163:**389-399.

Heal, O. W. 1964. Observations on the seasonal and spatial-distribution of Testacea (Protozoa, Rhizopoda) in *Sphagnum*. *Journal of Animal Ecology,* **33:**395-412.

Heger, T. J., Mitchell, E. A. D. & Leander, B. S. 2013. Holarctic phylogeography of the testate amoeba Hyalosphenia papilio (Amoebozoa: Arcellinida) reveals extensive genetic diversity explained more by environment than dispersal limitation. *Molecular Ecology,* **22:**5172-5184.

Heger, T. J., Mitchell, E. A. D., Todorov, M., Golemansky, V., Lara, E., Leander, B. S. & Pawlowski, J. 2010. Molecular phylogeny of euglyphid testate amoebae (Cercozoa: Euglyphida) suggests transitions between marine supralittoral and freshwater/terrestrial environments are infrequent. *Molecular Phylogenetics and Evolution,* **55:**113-122.

Huang, J. & Katz, L. A. 2014. Nanochromosome Copy Number Does not Correlate with RNA Levels Though Patterns are Conserved between Strains of the Ciliate Morphospecies *Chilodonella uncinata*. *Protist,* **165:**445-451.

Jassey, V. E. J., Chiapusio, G., Mitchell, E. A. D., Binet, P., Toussaint, M. L. & Gilbert, D. 2011. Fine-Scale Horizontal and Vertical Micro-distribution Patterns of Testate Amoebae Along a Narrow Fen/Bog Gradient. *Microbial Ecology,* **61:**374-385.

Jassey, V. E. J., Chiapusio, G., Binet, P., Buttler, A., Laggoun-Defarge, F., Delarue, F., Bernard, N., Mitchell, E. A. D., Toussaint, M. L., Francez, A. J. & Gilbert, D. 2013. Above- and belowground linkages in Sphagnum peatland: climate warming affects plant-microbial interactions. *Global Change Biology,* **19:**811-823.

Jung, J.-H., Kim, S., Ryu, S., Kim, M.-S., Baek, Y.-S., Kim, S.-J., Choi, J.-K., Park, J.-K. & Min, G.-S. 2012. Development of single-nucleotide polymorphism-based phylum-specific PCR amplification technique: application to the community analysis using ciliates as a reference organism. *Molecules and cells***:**1-9.

Katoh, K. & Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.,* **30:**772-780.

Klappenbach, J. A., Saxman, P. R., Cole, J. R. & Schmidt, T. M. 2001. rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Res,* **29:**181-4.

Kosakyan, A., Gomaa, F., Lara, E. & Lahr, D. J. G. 2016. Current and future perspectives on the systematics, taxonomy and nomenclature of testate amoebae. *European Journal of Protistology,* **55:**105-117.

Kosakyan, A., Gomaa, F., Mitchell, E. A. D., Heger, T. J. & Lara, E. 2013. Using DNA-barcoding for sorting out protist species complexes: A case study of the Nebela tincta-collaris-bohemica group (Amoebozoa; Arcellinida, Hyalospheniidae). *European Journal of Protistology,* **49:**222-237.

Kosakyan, A., Heger, T. J., Leander, B. S., Todorov, M., Mitchell, E. A. & Lara, E. 2012. COI barcoding of Nebelid testate amoebae (Amoebozoa: Arcellinida): extensive cryptic diversity and redefinition of the Hyalospheniidae Schultze. *Protist,* **163:**415-34.

Lahr, D. J., Bosak, T., Lara, E. & Mitchell, E. A. 2015. The Phanerozoic diversification of silica-cycling testate amoebae and its possible links to changes in terrestrial ecosystems. *PeerJ,* **3:**e1234.

Lahr, D. J. G. & Lopes, S. G. B. C. 2006. Morphology, biometry, ecology and biogeography of five species of Difflugia Leclerc, 1815 (Arcellinida : Difflugiidae), from Tiete River, Brazil. *Acta Protozoologica,* **45:**77-90.

Lahr, D. J. G. & Katz, L. A. 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques,* **47:**857-863.

Lahr, D. J. G., Grant, J., Nguyen, T., Lin, J. H. & Katz, L. A. 2011. Comprehensive phylogenetic reconstruction of Amoebozoa based on concatenated analyses of SSU-rDNA and actin genes. *Plos One,* **6**.

Lamentowicz, M. & Mitchell, E. A. 2005. The ecology of testate amoebae (Protists) in sphagnum in North-western Poland in relation to peatland ecology. *Microbial Ecology,* **50:**48-63.

Lamentowicz, M., Bragazza, L., Buttler, A., Jassey, V. E. J. & Mitchell, E. A. D. 2013. Seasonal patterns of testate amoeba diversity, community structure and species-environment relationships in four Sphagnum-dominated peatlands along a 1300 m altitudinal gradient in Switzerland. *Soil Biology & Biochemistry,* **67:**1-11.

Lara, E., Heger, T. J., Ekelund, F., Lamentowicz, M. & Mitchell, E. A. D. 2008. Ribosomal RNA genes challenge the monophyly of the hyalospheniidae (Amoebozoa : Arcellinida). *Protist,* **159:**165-176.

Lozupone, C. & Knight, R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology,* **71:**8228-8235.

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J. & Knight, R. 2011. UniFrac: an effective distance metric for microbial community comparison. *The ISME journal,* **5:**169.

Maddison, W. P. & Maddison, D. R. (2011) Mesquite: a modular system for evolutionary analysis. Version 2.75 ed.

Mahe, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. 2014. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ,* **2:**e593.

Mahe, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ,* **3:**e1420.

Mahe, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Sernaker, S., Siemensmeyer, T., Trautmann, I., Romac, S., Berney, C., Kozlov, A., Mitchell, E. A. D., Seppey, C. V. W., Egge, E., Lentendu, G., Wirth, R., Trueba, G. & Dunthorn, M. 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat Ecol Evol,* **1:**91.

Maki, A., Salmi, P., Mikkonen, A., Kremp, A. & Tiirola, M. 2017. Sample Preservation, DNA or RNA Extraction and Data Analysis for High-Throughput Phytoplankton Community Sequencing. *Frontiers in Microbiology,* **8**.

Markel, E. R., Booth, R. K. & Qin, Y. M. 2010. Testate amoebae and delta C-13 of Sphagnum as surface-moisture proxies in Alaskan peatlands. *Holocene,* **20:**463-475.

McKeown, M. M., Wilmshursts, J. M., Duckert, C., Woods, J. R. & Mitchell, E. A. D. 2019. Assessing the ecological value of small testate amoebae (<45 mu m) in New Zealand peatlands. *European Journal of Protistology,* **68:**1-16.

McMurdie, P. J. & Holmes, S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one,* **8:**e61217.

Meisterfeld, R. 2002. Order Arcellinida Kent, 1880. *In*: Lee, J. J., F., L. G. & P., B. (eds.) The Illustrated Guide to the Protozoa. 2nd ed. Allen Press, Lawrence, Kansas, USA.827-860.

Meisterfeld, R. 1977. Horizontal and vertical distribution of testacea (Rhizopoda-testacea) in Sphagnum. *Archiv Fur Hydrobiologie,* **79:**319-356.

Mignot, J.-P. & Raikov, I. B. 1992. Evidence for meiosis in the testate amoeba *Arcella*. *Journal of Eurkaryotic Microbiology,* **39:**287-289.

Mitchell, B., Jacobs, R., Li, J., Chien, S. & Kintner, C. 2007. A positive feedback mechanism governs the polarity and motion of motile cilia. *Nature,* **447:**97-U8.

Mitchell, E. A. D., Charman, D. J. & Warner, B. G. 2008. Testate amoebae analysis in ecological and paleoecological studies of wetlands: past, present and future. *Biodiversity and Conservation,* **17:**2115-2137.

Mitchell, E. A. D., Borcard, D., Buttler, A. J., Grosvernier, P., Gilbert, D. & Gobat, J. M. 2000. Horizontal distribution patterns of testate amoebae (Protozoa) in a Sphagnum magellanicum carpet. *Microbial Ecology,* **39:**290-300.

Molin, S. & Givskov, M. 1999. Application of molecular tools for in situ monitoring of bacterial growth activity. *Environ Microbiol,* **1:**383-91.

Ogden, C. G. & Hedley, R. H. 1980. An atlas of freshwater testate amoebae. Oxford University Press for the British Museum (Natural History), Oxford. p. 222 p.

Oksanen, J., Blanchet, F., Kindt, R., Legendre, P., Minchin, P. & O'hara, R. (2016) vegan: Community Ecology Package. R package version 2.3-4 [WWW document].

Oliverio, A., Lahr, D. J. G., Nguyen, T. & Katz, L. A. 2014. Cryptic diversity within morphospecies of testate (shelled) amoebae in New England bogs and fens. *Protist,* **165:**196-207.

Patterson, R. T., Lamoureux, E. D., Neville, L. A. & Macumber, A. L. 2013. Arcellacea (testate lobose amoebae) as pH indicators in a pyrite mine-acidified lake, Northeastern Ontario, Canada. *Microb Ecol,* **65:**541-54.

Payne, R. J. 2013. Seven Reasons Why Protists Make Useful Bioindicators. *Acta Protozoologica,* **52:**105-113.

Porfirio-Sousa, A. L., Ribeiro, G. M. & Lahr, D. J. G. 2017. Morphometric and genetic analysis of *Arcella intermedia* and *Arcella intermedia laevis* (Amoebozoa, Arcellinida) illuminate phenotypic plasticity in microbial eukaryotes. *European Journal of Protistology,* **58:**187-194.

Porter, S. M. & Knoll, A. H. 2000. Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. *Paleobiology,* **26:**360-385.

Price, M. N., Dehal, P. S. & Arkin, A. P. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. *Plos One,* **5:**e9490.

Price, M. N., Dehal, P. S. & Arkin, A. P. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution,* **26:**1641-50.

Rice, P., Longden, I. & Bleasby, A. 2000. EMBOSS: The European molecular biology open software suite. *Trends in Genetics,* **16:**276-277.

Roe, H. M. & Patterson, R. T. 2014. Arcellacea (testate amoebae) as bio-indicators of road salt contamination in lakes. *Microb Ecol,* **68:**299-313.

Roe, H. M., Elliott, S. M. & Patterson, R. T. 2017. Re-assessing the vertical distribution of testate amoeba communities in surface peats: Implications for palaeohydrological studies. *European Journal of Protistology,* **60:**13-27.

Schmidt, T. M., DeLong, E. F. & Pace, N. R. 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol,* **173:**4371-8.

Schonborn, W. 1962. Ecology of Testacea on Sphagnum bag, mass and earth. *Limnologica,* **1:**231-254.

Singer, D., Kosakyan, A., Pillonel, A., Mitchell, E. A. D. & Lara, E. 2015. Eight species in the Nebela collaris complex: Nebela gimlii (Arcellinida, Hyalospheniidae), a new species described from a Swiss raised bog. *European Journal of Protistology,* **51:**79-85.

Singer, D., Mitchell, E. A. D., Payne, R. J., Blandenier, Q., Duckert, C., Fernndez, L. D., Fournier, B., Hernndez, C. E., Granath, G., Rydin, H., Bragazza, L., Koronatova, N. G., Goia, I., Harris, L. I., Kajukalo, K., Kosakyan, A., Lamentowicz, M., Kosykh, N. P., Vellak, K. & Lara, E. 2019. Dispersal limitations and historical factors determine the biogeography of specialized terrestrial protists. *Molecular Ecology,* **28:**3089-3100.

Sisson, C., Gulla-Devaney, B., Katz, L. A. & Grattepanche, J. D. 2018. Seed bank and seasonal patterns of the eukaryotic SAR (Stramenopila, Alveolata and Rhizaria) clade in a New England vernal pool. *Journal of Plankton Research,* **40:**376-390.

Smith, H. G., Bobrov, A. & Lara, E. 2008. Diversity and biogeography of testate amoebae. *Biodiversity and Conservation,* **17:**329-343.

Song, L. H., Gilbert, D. & Wu, D. H. 2016. Vertical micro-distribution of microbial communities living in *Sphagnum fallax*. *Aquatic Microbial Ecology,* **77:**1-10.

Song, L. H., Li, H. K., Wang, K. H., Yan, X. M. & Wu, D. H. 2018. Seasonal dynamics in the community structure and trophic structure of testate amoebae inhabiting the Sanjiang peatlands, Northeast China. *European Journal of Protistology,* **63:**51-61.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics,* **30:**1312-1313.

Steele, R. E., Nasser, N. A., Patterson, R. T., Gregory, B. R. B., Roe, H. M. & Reinhardt, E. G. 2018. An Assessment of Sub-Meter Scale Spatial Variability of Arcellinida (Testate Lobose Amoebae) Assemblages in a Temperate Lake: Implications for Limnological Studies. *Microbial Ecology,* **76:**680-694.

Tolonen, K., Warner, B. G. & Vasander, H. 1994. Ecology of Testaceans (Protozoa, Rhizopoda) in mires in southern Finland. 2. Multivariate-analysis. *Archiv Für Protistenkunde,* **144:**97-112.

Tolonen, K., Warner, B. G. & Vasander, H. 1992. Ecology of Testaceans (Protozoa, Rhizopoda) in mires in southern Finland.1. Autoecology. *Archiv Für Protistenkunde,* **142:**119-138.

Turner, T. E. & Swindles, G. T. 2012. Ecology of Testate Amoebae in Moorland with a Complex Fire History: Implications for Ecosystem Monitoring and Sustainable Land Management. *Protist,* **163:**844-855.

Volkova, E. & Smirnov, A. 2016. Regeneration of test in testate amoebae of the genus *Arcella* (Tubulinea, Arcellinida). *European Journal of Protistology,* **55:**128-140.

Wang, C., Zhang, T., Wang, Y., Katz, L. A., Gao, F. & Song, W. 2017. Disentangling sources of variation in SSU rDNA sequences from single cell analyses of ciliates: impact of copy number variation and experimental error. *Proc Biol Sci,* **284**.

Wang, Y. R., Wang, C. D., Jiang, Y. H., Katz, L. A., Gao, F. & Yan, Y. 2019. Further analyses of variation of ribosome DNA copy number and polymorphism in ciliates provide insights relevant to studies of both molecular ecology and phylogeny. *Science China-Life Sciences,* **62:**203-214.

Warner, B. G., Asada, T. & Quinn, N. P. 2007. Seasonal influences on the ecology of testate amoebae (Protozoa) in a small Sphagnum peatland in southern Ontario, Canada. *Microb Ecol,* **54:**91-100.

Wilkinson, D. M. 2008. Testate amoebae and nutrient cycling: peering into the black box of soil ecology. *Trends in Ecology & Evolution,* **23:**596-599.

Worden, A. Z. & Binder, B. J. 2003. Growth regulation of rRNA content in Prochlorococcus and Synechococcus (marine cyanobacteria) measured by whole-cell hybridization of rRNA-targeted peptide nucleic acids. *Journal of Phycology,* **39:**527-534.

Yeates, G. W. & Foissner, W. 1995. Testate amebas as predators of nematodes. *Biology and Fertility of Soils,* **20:**1-7.

Yu, Z., Beilman, D. W., Frolking, S., MacDonald, G. M., Roulet, N. T., Camill, P. & Charman, D. J. 2011. Peatlands and their role in the global carbon cycle. *EOS, Transactions, American Geophysical Union,* **92:**97-109.

FIGURE LEGENDS:

Fig. 1: Considerable diversity among the 92 focal Arcellinida lineages amplified with our newly-designed primers. Unlabeled branches are OTUs, with their numbers in the first column. Major clades of sequences are labeled, and unlabeled regions are either taxa represented by single sequences (e.g. OTU525 is identical to *Argynnia dentistoma* sequence) or are lineages for which taxonomy cannot be readily assigned (See Table 1). The three columns are read number (>100 = yellow, >1,000 = orange, >10,000 red), abundance in surface (green) and/or cores (brown) in Experiment 1; and distribution among size classes (e.g. 2-10 µm, 10-35 µm, 35-100 µm, 100-300µm) for RNAs in Experiment 2 (see methods for more details). For latter two columns, absence of notation means OTU was rare or absent in this experiment. The names used here match those in individual sequence files, and we refer the reader to discussion in text on changing taxonomy in this field.

Fig. 2: Principal coordinates analysis of combined samples from Experiment 1 and 2 loosely separates communities into categories: surface RNAs (Experiment 1 = light green; Experiment 2 = dark green) above the line and core RNAs (Experiment 1 = light blue) and surface DNAs (Experiment 2 = dark blue). Analyses is based on presence/absence and uses UniFrac distances to account for evolutionary relatedness (see methods). For more sample details see Fig. S1.

Fig. 3: Principal coordinates analysis of RNA samples from Experiment 1 separate communities at the surface (open green circles) from those in the two cores (closed circles). The two core transects (C1 and C2) are more distinct from each other than the surface transects (T1 and T2), though there is no clear separation by segment (e.g.0-6 cm, 7-12 cm, etc.). Other notes as in Fig. 2.

Fig. 4: Principal coordinates analysis of RNA samples from Experiment 2 separate communities based on size (Axis 1), with the largest communities (100-300um = red dots) being most distinct. Along the y axis as all of the station 1 (S1; those closest to woodland edge of fen) communities are distinct from other communities. T = transects, S = station. Other notes as in Fig. 2

SUPPLEMENTAL MATERIALS:

Table S1: GenBank sequences of Arcellinida and other Amoebozoa used for primer design

Table S2: Arcellinida-specific primers for SSU-rDNA and COI, their sequences and performance

Table S3: Details on application of primer pairs

Table S4: Table of all 317 OTUs, their closest matches with GenBank data and their abundance in the different samples

Table S5: The 92 OTUs that were the focus of the distribution analyses

Table S6: Results of the morphospecies counts

Figure S1: Principal coordinates analysis of combined samples of Experiments 1 and 2 and the grouping of DNA and RNA samples

File S1: Amoebozoa sequence database from GenBank

File S2: Outgroup sequence database from GenBank

File S3: Alignment of Amoebozoa and outgroup sequences

File S4: Constrained tree

File S5: Alignment of Amoebozoa sequences

File S6: Phylogeny of OTUs plus Amoebozoa plus outgroups

File S7: Phylogeny of Amoebozoa OTUs (from file S6) plus Amoebozoa

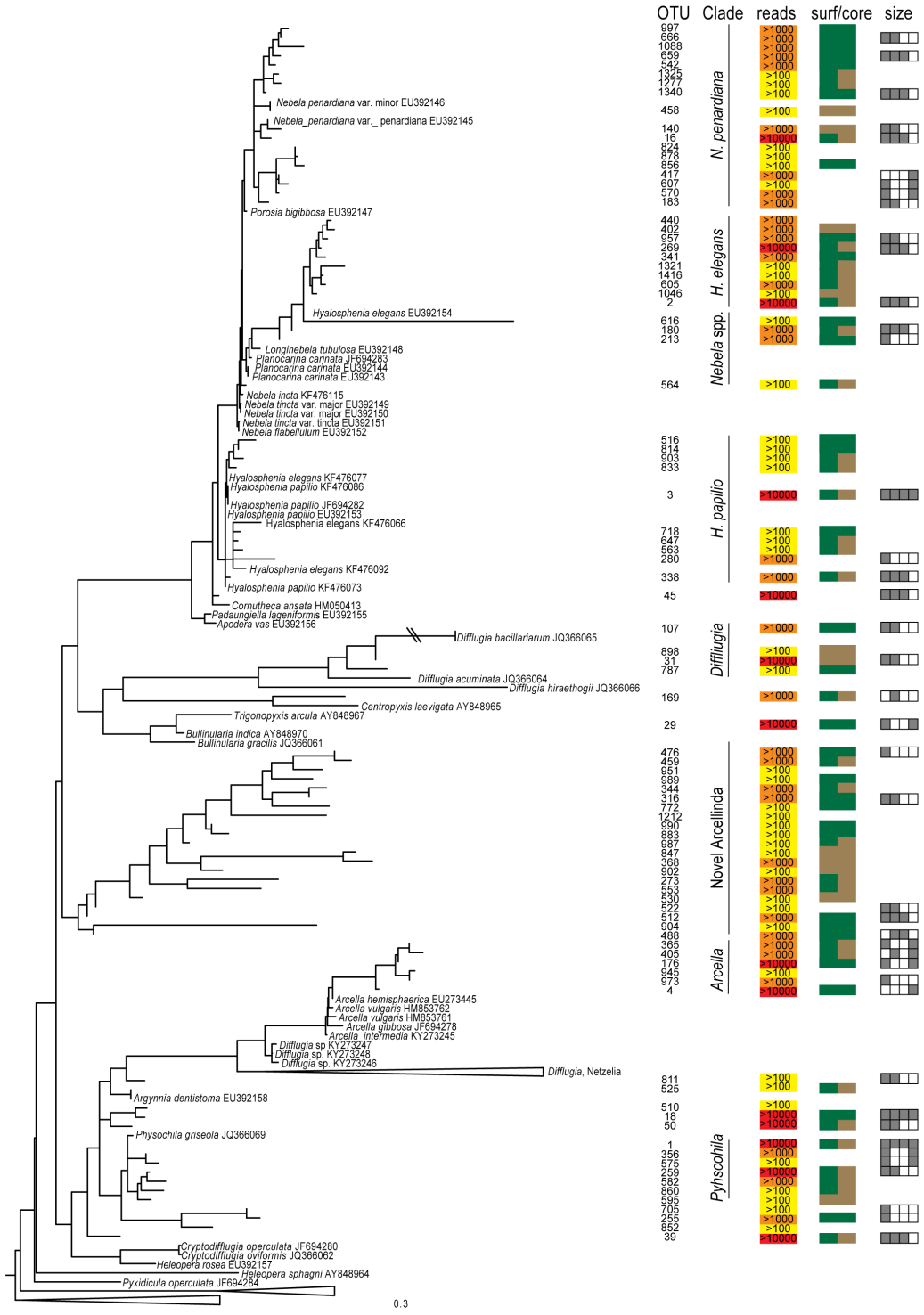
OTU
Clade
reads
surf/core
size
Nebela penardiana var. minor EU392146
Nebela_penardiana var._ penardiana EU392145
Porosia bigibbosa EU392147
Hyalosphenia elegans EU392154
Longinebela tubulosa EU392148
Planocarina carinata JF694283
Planocarina carinata EU392144
Planocarina carinata EU392143
Nebela incta KF476115
Nebela tincta var. major EU392149
Nebela tincta var. major EU392150
Nebela tincta var. tincta EU392151
Nebela flabellulum EU392152
Hyalosphenia elegans KF476077
Hyalosphenia papilio KF476086
Hyalosphenia papilio JF694282
Hyalosphenia papilio EU392153
Hyalosphenia elegans KF476066
Hyalosphenia elegans KF476092
Hyalosphenia papilio KF476073
Cornutheca ansata HM050413
Padaungiella lageniformis EU392155
Apodera vas EU392156
Difflugia bacillariarum JQ366065
Difflugia acuminata JQ366064
Difflugia hiraethogii JQ366066
Centropyxis laevigata AY848965
Trigonopyxis arcula AY848967
Bullinularia indica AY848970
Bullinularia gracilis JQ366061
Arcella hemisphaerica EU273445
Arcella vulgaris HM853762
Arcella vulgaris HM853761
Arcella gibbosa JF694278
Arcella_intermedia KY273245
Difflugia sp KY273247
Difflugia sp. KY273248
Difflugia sp. KY273246
Difflugia, Netzelia
Argynnia dentistoma EU392158
Physochila griseola JQ366069
Cryptodifflugia operculata JF694280
Cryptodifflugia oviformis JQ366062
Heleopera rosea EU392157
Heleopera sphagni AY848964
Pyxidicula operculata JF694284
0.3
N. penardiana
H. elegans
Nebela spp.
H. papilio
Difflugia
Novel Arcellinda
Arcella
Pyhscohila

Fig. 1

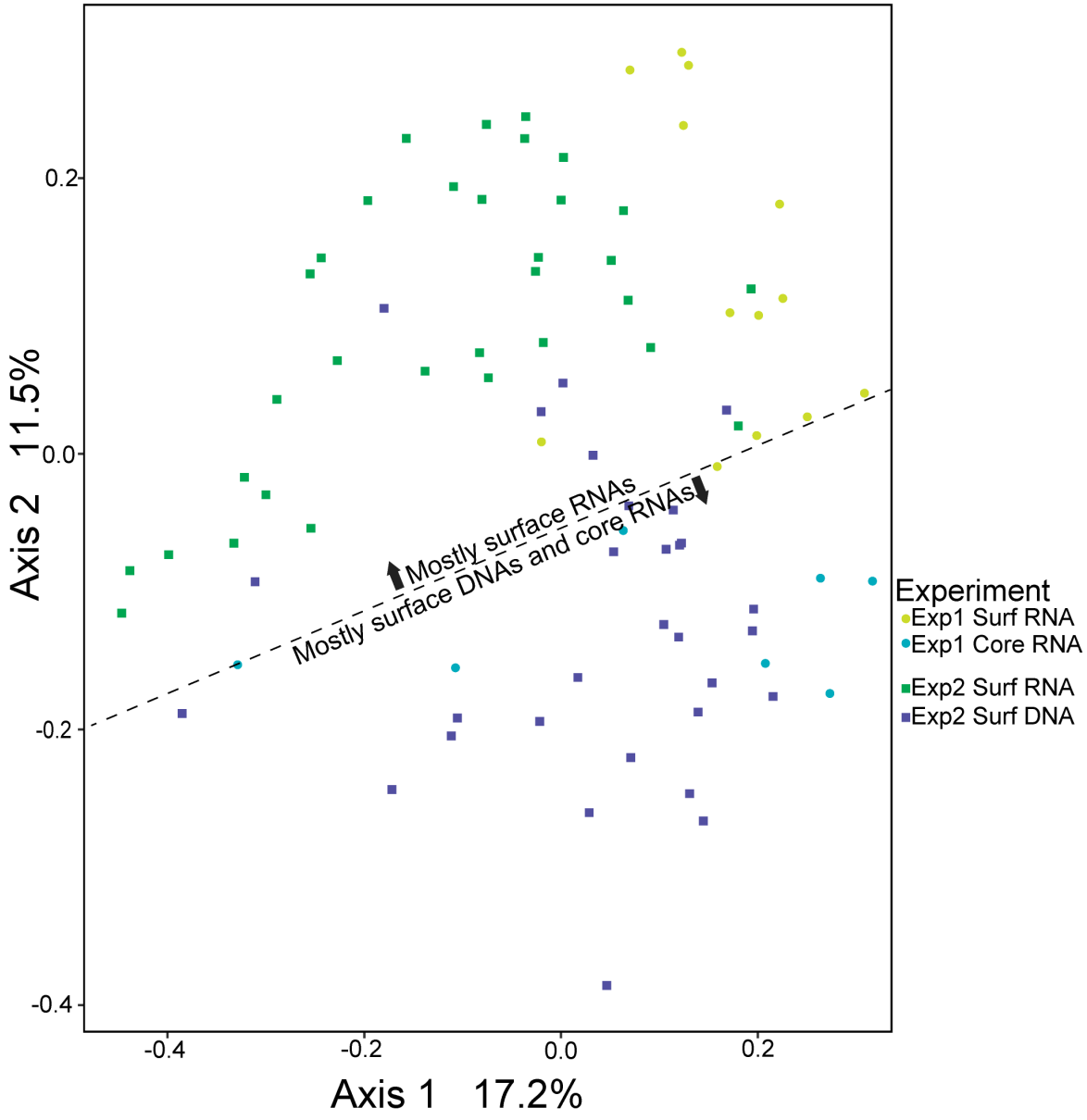
0.2
0.0
-0.2
-0.4
Axis 2 11.5%
-0.4
-0.2
0.0
0.2
Axis 1 17.2%
Mostly surface RNAs
Mostly surface DNAs and core RNAs
Experiment
Exp1 Surf RNA
Exp1 Core RNA
Exp2 Surf RNA
Exp2 Surf DNA

Fig. 2

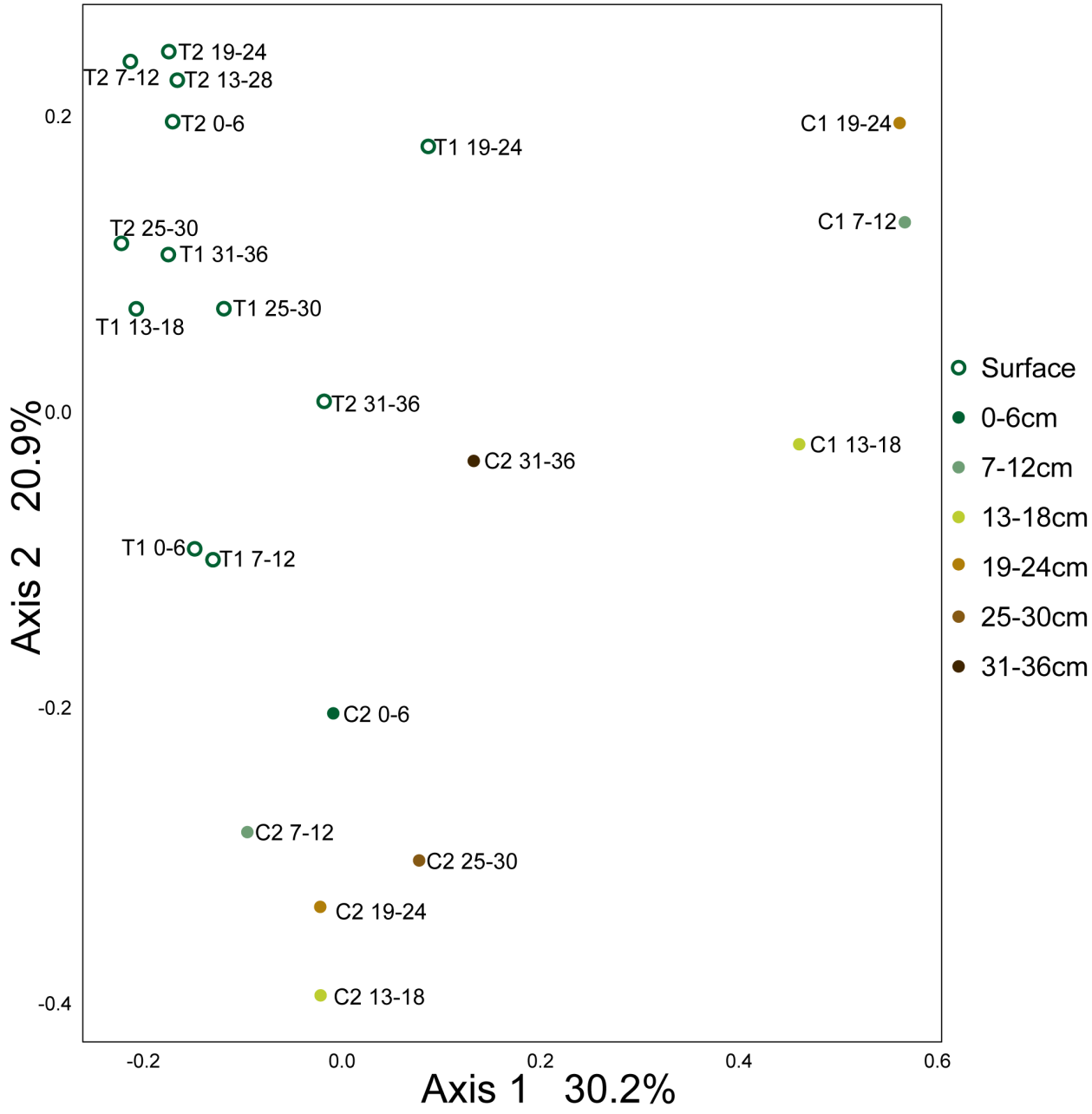
T2 19-24
T2 7-12
T2 13-28
T2 0-6
T1 19-24
C1 19-24
C1 7-12
T2 25-30
T1 31-36
T1 25-30
T1 13-18
T2 31-36
C1 13-18
C2 31-36
T1 0-6
T1 7-12
C2 0-6
C2 7-12
C2 25-30
C2 19-24
C2 13-18
0.2
0.0
-0.2
-0.4
Axis 2 20.9%
-0.2
0.0
0.2
0.4
0.6
Axis 1 30.2%
Surface
0-6cm
7-12cm
13-18cm
19-24cm
25-30cm
31-36cm

Fig. 3

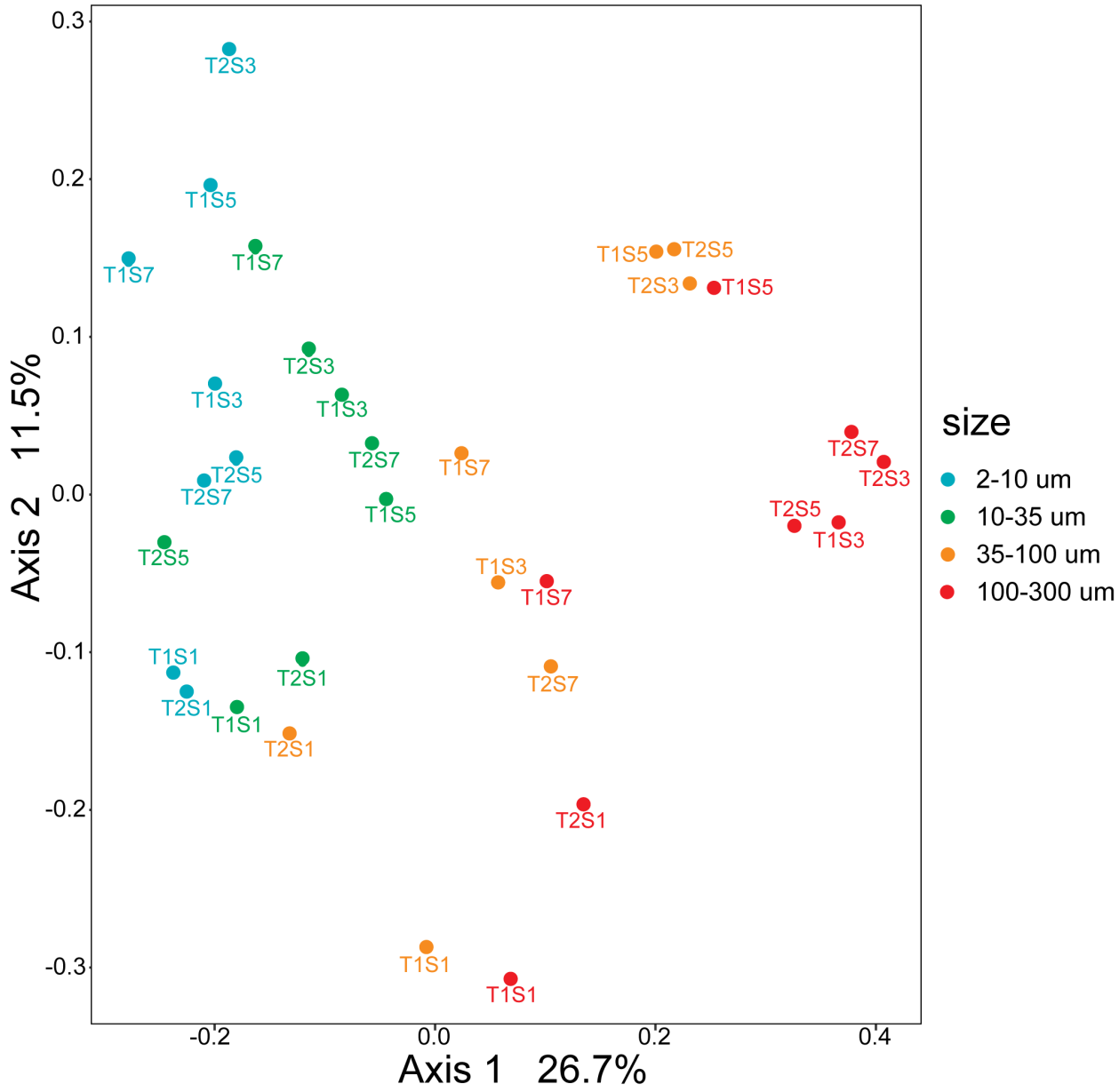
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
Axis 2 11.5%
-0.2
0.0
0.2
0.4
Axis 1 26.7%
T2S3
T1S5
T1S7
T1S7
T2S3
T1S3
T1S3
T2S7
T2S5
T2S7
T1S7
T1S5
T2S5
T1S3
T1S7
T1S1
T2S1
T2S1
T1S1
T2S1
T2S7
T2S1
T1S1
T1S1
T1S5
T2S5
T2S3
T1S5
T2S7
T2S3
T2S5
T1S3
size
2-10 um
10-35 um
35-100 um
100-300 um

Fig. 4