RESEARCH ARTICLE

Phylogenetic relatedness drives protist assembly in marine and terrestrial environments

Guillaume Lentendu[1,2,3] (iD) | Micah Dunthorn[4,5,6] (iD)

[1]Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

[2]Department of Ecology, University of Kaiserslautern, Kaiserslautern, Germany

[3]technische universitaet kaiserslautern, Erwin-Schrödinger-Straße 52, 67663, Kaiserslautern, Germany

[4]Natural History Museum, University of Oslo, Oslo, Norway

[5]Eukaryotic Microbiology, Faculty of Biology, University of Duisburg-Essen, Essen, Germany

[6]Centre for Water and Environmental Research (ZWU), University of Duisburg-Essen, Essen, Germany

Correspondence
Guillaume Lentendu, Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland.
Email: guillaume.lentendu@unine.ch

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: DU1319/5-1

Editor: Janne Soininen

Abstract

Aim: Assembly of protist communities is known to be driven mainly by environmental filtering, but the imprint of phylogenetic relatedness is unknown. In this study, we aimed to test the degree to which co-occurrences and co-exclusions of protists in different phylogenetic relatedness classes deviate from random expectation in two ecosystems, in order to link them to ecological processes.

Location: Global open oceans and Neotropical rain-forest soils.

Major taxa: Protists.

Time period: 2009–2013.

Methods: Protist metabarcoding data originated from two large-scale studies. Co-occurrence and co-exclusion networks were constructed using a recent method combining a null distribution model with Spearman's rank correlation coefficients among pairs of operational taxonomic units. Phylogenetic relatedness was estimated using either global pairwise sequence distance or phylogenetic distance inferred from best maximum-likelihood trees derived from multiple alignments of operational taxonomic unit representative sequences. The significance of observed patterns relating networks and phylogenies was evaluated by distance classes against two null models, in which either the tips of the phylogenetic trees or the network edges were randomized.

Results: Closely related protists co-occurred more often than expected by chance in all datasets, but also co-excluded less often than expected by chance in the marine dataset only. Concurrent excesses of co-occurrences and co-exclusions were observed at intermediate phylogenetic distances in the marine dataset.

Main conclusions: This suggests that environmental filtering and dispersal limitation are the dominant forces driving protist co-occurrences in both environments, whereas a signal of competitive exclusion is detected only in the marine environment. Differences in co-exclusion are potentially linked to the individual environments, in that marine waters are more homogeneous, whereas rain-forest soils contain a myriad of nutrient-rich microenvironments, reducing the strength of mutual exclusion.

KEYWORDS
co-exclusion, competitive exclusion, co-occurrence, environmental filtering, macroecology, network, null model, phylogenetic diversity, protists

1 | INTRODUCTION

There is a long history of research trying to elucidate why species are present in a specific environment and why multiple species are found together (Darwin, 1859; Gause, 1934; Humboldt & Bonpland, 1805). Species sharing the same ecological niche tend to co-occur owing to environmental filtering and dispersal limitation. Closely related species are more likely to co-occur owing to their shared evolutionary history (e.g., common ancestor, shared traits) and their potential limited dispersal and establishment abilities. These processes can be balanced by density-dependent negative biotic interactions, such as competitive exclusion when functionally similar species require the same resource and co-exclude themselves. These deterministic processes as defined by the niche theory can then be opposed to stochastic processes as defined by the neutral theory, and together form the different forces driving community assembly (Hubbell, 2001; Kraft et al., 2015; Vandermeer, 1972). Environmental filtering and dispersal limitation have been identified as the main drivers shaping the assembly of most protists in different environments (Boenigk et al., 2018; de Vargas et al., 2015; del Campo et al., 2015; Lentendu et al., 2018; Mahé et al., 2017; Singer et al., 2018; Wetzel et al., 2012), whereas competition has been tested formally in laboratory conditions only (Saleem et al., 2012; Violle et al., 2011). These two processes with opposing effects on community assembly have already been observed concurrently in marine ciliates, but they operate at different spatial scales, with competitive exclusion being detected only at the local scale, whereas environmental filtering is detected only at the regional scale, and they have not been observed globally (Azovsky, 1996).

Environmental filtering is a recent concept, first postulated as the "coexistence principle" (Den Boer, 1980). In environmental microbial ecology, environmental filtering is often considered as the prevalent limiting parameter of species occurrence (Khomich et al., 2017; Lauber et al., 2008; Lentendu et al., 2018; Philippot et al., 2010; Singer et al., 2018; Tedersoo et al., 2016; Weißbecker et al., 2018; Zinger et al., 2011) and is linked directly to the ecological niche of microbes. The ecological niche of microbes is difficult to measure without cultivation (Lennon et al., 2012; Martiny et al., 2015); therefore, function and functional similarity are deduced from either taxonomic or phylogenetic similarity of recovered sequences in large-scale studies. Environmental filtering is inferred from the non-random co-occurrence of members of a taxon or a clade or from a clade or taxon occurring in a restricted set of habitats. Thus, environmental filtering, when analysed in a phylogenetic context, often assumes phylogenetic niche conservatism; that is, the long-term retention of ecological traits among closely related species (Wiens et al., 2010). Phylogenetic niche conservatism was shown in bacteria, mainly for complex functional traits that are conserved inside single clades (Martiny et al., 2013). Under phylogenetic niche conservatism, evolutionarily close species are more likely to share the same ecological niche; hence, they tend to be filtered into the same habitats. With

this assumption, environmental filtering can be tested using measures of phylogenetic divergence (e.g., mean pairwise distance, mean nearest taxon distance; but see Tucker et al., 2017), with phylogenetic overclustering (i.e., low phylogenetic divergence) being interpreted as sign of environmental filtering. This sample-wide approach has been used to support environmental filtering of trees, bacteria and protists along habitat and nutrient gradients (Horner-Devine & Bohannan, 2006; Kembel & Hubbell, 2006; Singer et al., 2018). However, most studies considering environmental filtering do not account for biotic interactions, which could produce similar results (Kraft et al., 2015).

Competition has long been known experimentally, and it was hypothesized to drive co-exclusion in an initial experimental study involving protists (Gause, 1934). Competitive exclusion was initially viewed as an evolutionary pressure triggering trait divergence of related species, allowing them to escape competition and to persist in the same habitat, as originally observed for Darwin's finches (Darwin, 1859). This assumption was further formalized with the phylogenetic limiting similarity hypothesis, in which phylogenetically related species compete more strongly owing to niche overlap, thereby limiting the number of related species that can coexist (Macarthur & Levins, 1967). By assuming phylogenetic niche conservatism, it is expected that competitive exclusion will affect only closely related species, meaning that phylogenetic overdispersion (i.e., high phylogenetic divergence) of natural communities is interpreted as a sign of competitive exclusion. This approach has allowed for the identification of one tree family presenting signs of competitive exclusion in a tropical forest (Manel et al., 2014). Competition does not necessarily lead to exclusion when, for example, competition is symmetrical or when other biotic interactions (e.g., mutualism or herbivory) reduce or neutralize the competition (Lamb & Cahill, 2008; Müller et al., 2012; Olff & Ritchie, 1998). Further experimental evidence has shown that for protist species, competition will lead to exclusion more quickly when species are phylogenetically related, with a direct relationship to phylogenetically conserved traits (e.g., mouth size; Violle et al., 2011). The "paradox of the plankton" was also considered to be an opposite example of competitive exclusion, with the coexistence of high numbers of species using the same resources (Hutchinson, 1961). It was shown that this pattern is explained by the competition itself, which leads to only short-term exclusion in a system that never reaches an equilibrium (Huisman & Weissing, 1999). In plant ecology, studies measuring the strength of competition have shown that, depending on clades or depending on soil conditions, there will be more or less competition between related species, meaning that no generalization of the "competition-relatedness" hypothesis is possible (Burns & Strauss, 2011; Cahill et al., 2008). The exclusion of closely related species owing to competition can thus be viewed as a special case of the coexistence theory (Mayfield & Levine, 2010). Only a few large-scale studies have tested for phylogenetic overdispersion and exclusion patterns in protistan species (Azovsky, 1996).

In today's very large environmental sequencing datasets, microbial taxa are characterized using operational taxonomic

units (OTUs), which are used as a proxy for molecular species (Santoferrara et al., 2020). At the same time, co-occurrence and co-exclusion network analyses have become standard in environmental microbial ecology, with a predominance of studies interested in co-occurrence patterns among and between taxonomic groups with a presumed function (Chow et al., 2014; Lima-Mendez et al., 2015; Milici et al., 2016; Steele et al., 2011). In contrast to phylogenetic divergence analyses conducted at the sample level, co-occurrence and co-exclusion network analyses allow the extraction of statistically significant pairs of co-occurring/co-excluding OTUs at the whole-study level. By comparing observed co-occurrences with random co-occurrences among the regional pool of OTUs, signals for potential biotic interactions such as parasitism, predation or viral infection have been disclosed (Lentendu et al., 2014; Lima-Mendez et al., 2015; Steele et al., 2011). By taking advantage of the modular structure of the co-occurrence networks, microbial occurrences have also been linked to habitat preference, which can be interpreted as the signal for environmental filtering (de Menezes et al., 2015; Lentendu et al., 2014; Milici et al., 2016; Morriën et al., 2017). However, studies have yet to integrate phylogenetic relatedness as an explanatory parameter for network structure, which has been introduced recently for soil bacteria (Goberna et al., 2019).

Here, we describe a new analytical approach to evaluate community assembly processes by decomposing the co-occurrence and co-exclusion networks among phylogenetic relatedness classes. By looking at the excess or deficit of co-occurrence or co-exclusion in a class of organism with increasing phylogenetic relatedness, we can test the possible assembly mechanisms in natural protistan communities. Under the assumption of phylogenetic niche conservatism, we test the following hypotheses: (a) if environmental filtering dominates, phylogenetically related OTUs will co-occur more and co-exclude less often than expected by chance, and vice versa for pairs of OTUs with intermediate phylogenetic relatedness; and (b) if competitive exclusion dominates, phylogenetically related OTUs will co-occur less and co-exclude more often than expected by chance, and vice versa for pairs of OTUs with intermediate phylogenetic relatedness. To evaluate these hypotheses, we use two of the largest environmental sequencing protist datasets to date, namely the global open oceans from the study by de Vargas et al. (2015) and Neotropical rain-forest soils from the study by Mahé et al. (2017).

2 | MATERIALS AND METHODS

2.1 | Datasets

Protistan OTUs from the world's open oceans and seas came from the study by de Vargas et al. (2015). This marine dataset is composed of 355 samples collected at the surface and deep chlorophyll maximum (DCM) in six oceans and two seas, which produced 366,800,845 protist reads of the V9 hypervariable region of the small subunit ribosomal RNA (SSU rRNA) locus that clustered into 302,663 OTUs. To allow for comparison, the version of this marine dataset used here was re-analysed by Mahé et al. (2017). All filter-size-class libraries of either the surface or the DCM at a single station were pooled; hence, the number of samples used here was reduced to 47 for surface and 32 for DCM waters.

Protistan OTUs from three lowland Neotropical rain forests came from the study by Mahé et al. (2017). This terrestrial dataset is composed of 144 samples collected at the soil surface, which produced 46,652,206 protist reads of the V4 hypervariable region of the SSU rRNA locus that clustered into 26,860 OTUs. Sequence processing, OTU clustering with SWARM v.2 (Mahé et al., 2015), and taxonomic assignments using the PR^2 database (Guillou et al., 2013) are described in Mahé et al. (2017).

2.2 | Co-occurrence and co-exclusion networks

To infer protistan co-occurrences and co-exclusions from the marine and terrestrial datasets, networks were constructed using OTUs according to the method described by Connor et al. (2017). This method infers positive correlations (co-occurrences), which were expanded here also to infer negative correlations (co-exclusions; Supporting Information Supplementary Material). For that, correlation thresholds were determined for each dataset using communities with random distributions generated by OTU abundance swaps within each sample (i.e., fixed sample margin). This approach allows preservation of the observed distribution of Spearman's ρ in natural communities and is thus adequate to infer both significant positive and significant negative interactions compared to other random shuffling approaches (Supporting Information Figure S1). The resulting networks were composed of nodes (OTUs) that were connected by edges to one or more other nodes; these edges were either significant positive or significant negative correlations, which we interpret as instances of co-occurrences or co-exclusions, respectively. Analyses were restricted to OTUs occurring in > 30% of marine and > 10% of terrestrial samples, because OTUs with low occurrence cannot provide statistical significance (Connor et al., 2017).

2.3 | Pairwise sequence and phylogenetic distances

To infer the phylogenetic relatedness between the OTUs (nodes) in the networks, the representative sequences (the most abundant strictly identical amplicon) of each candidate OTU were used. These phylogenetic relatedness values between pairs of OTUs were then overlaid along the edges in the networks. Two methods were used to infer the phylogenetic relatedness. First, pairwise sequence distances were calculated using a Needleman–Wunsch approximation as implemented in SUMATRA v.1.0.34 (Mercier et al., 2013). This global pairwise sequence comparison did not account for any model of evolution. Second, phylogenetic distances were calculated by aligning the sequences using the FFT-NS-i strategy in MAFFT v.7.407 (Katoh

& Standley, 2013) and by finding the best maximum-likelihood tree using the GTRCAT model in RAxML v.8.2.12 (Stamatakis, 2014) with 128 random starting trees. The phylogenetic distance between each tree tip was then calculated with the "cophenetic" function in R (R Core Team, 2017).

2.4 | Null models

To infer the stochastic or deterministic origin of the associations between the networks (both co-occurrence and co-exclusion) and the phylogenetic relatedness, two null models were constructed. Null model 1 generated random phylogenetic relatedness values between nodes following Hardy (2008: model 1s) (Figure 1). These random values were made by shuffling the tip of the phylogenetic tree made of all candidate OTUs. The same random reordering of OTUs was applied to both pairwise sequence and phylogenetic distance matrices (i.e., reordering row and column names), and the distance value for each co-occurring or co-excluding OTU pair was extracted. The purpose of null model 1 was to test whether co-occurring or co-excluding OTUs were more or less phylogenetically related than expected by chance.

Null model 2 generated random edges between nodes following Chung and Lu (2002) (Figure 1). In these random networks, the total numbers of edges and nodes remained the same as in the observed network, but the number of edges from an individual node was drawn from a probability distribution in which edge probability depends on the cumulative observed degree of the two nodes involved. This null model produced networks with characteristics (e.g., modularity, diameter, clustering coefficient) more similar to natural networks compared with the most widely used Erdős–Rényi null model (Connor et al., 2017). The random networks were generated by the "sample_fitness" function in the R igraph package (Csardi & Nepusz, 2006). The purpose of null model 2 was to test whether

phylogenetically related OTUs co-occurred or co-excluded more or less than expected by chance.

2.5 | Statistical analyses

Null model constructions were repeated 1,000 times in order to test for a statistical difference with the observed data. Phylogenetic relatedness was aggregated in a step-wise manner, using a step of .01 for pairwise sequence distances and a step of .1 for phylogenetic distances. For each distance class, the number of co-occurring or co-excluding OTUs was accounted for in the observed and random networks, and a nonparametric p-value was calculated as the amount of time for which the observed number of co-occurrences or co-exclusions was higher or lower than in the null models. Differences between the observed networks and the null models were considered significant if the p-values were ≤ .05. Results were summarized for each distance class into the standardized effect size (SES), calculated according to Gotelli and McCabe (2002). By convention, an SES is considered as strong if it is at least two.

3 | RESULTS

3.1 | Network coverage

In order to test for a phylogenetic signal between co-occurring and co-exclusion OTUs with different phylogenetic relatedness, co-occurrence and co-exclusion networks were related to pairwise sequence and phylogenetic distances. Edges of connected OTUs in the networks were labelled with the phylogenetic relatedness distances, and the number of edges in each distance class were compared with the two null models. The marine protist networks consisted of 32–53% of candidate OTUs, whereas terrestrial protist networks
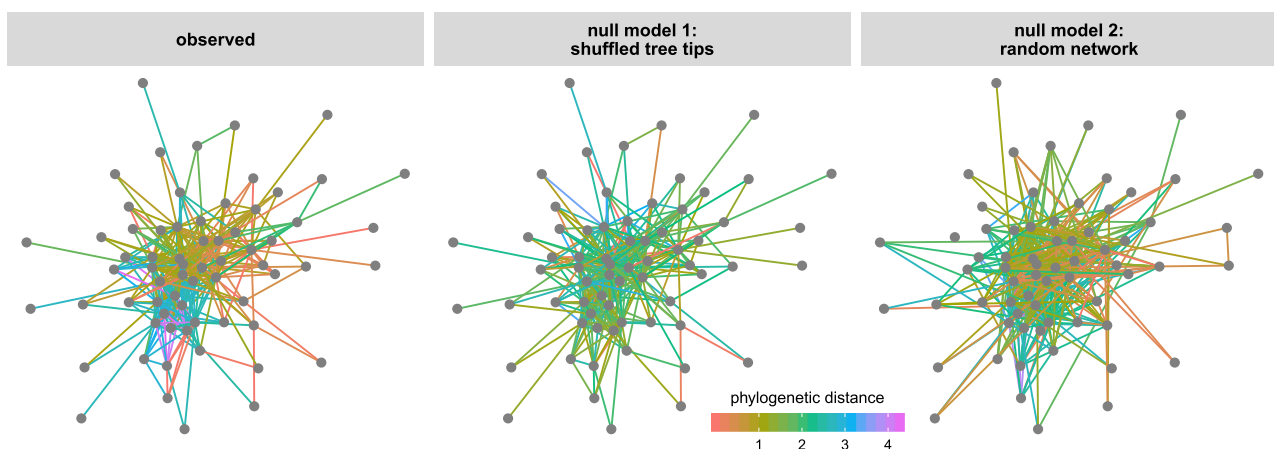

FIGURE 1 Null model effects on co-occurrence networks. Using the terrestrial protist co-occurrence network (observed), in which nodes are operational taxonomic units, edges are significant co-occurrences and edge colours phylogenetic distances. Null model 1 shuffled the tree tips, whereas null model 2 randomized the edges with a probability model (random network). The same approach was used for pairwise sequence distances and for co-exclusion networks [Colour figure can be viewed at wileyonlinelibrary.com]

included only 6–12% of candidate OTUs (Table 1). The network OTUs occurred in ≥ 32% of marine surface, ≥ 37% of marine DCM or ≥ 17% of terrestrial samples. The terrestrial co-exclusion network was much smaller than all the other networks.

The occurrence patterns of network OTUs were slightly skewed toward OTUs occurring in the highest number of samples and geographical units (i.e., seas, oceans and/or forests; Supporting Information Figure S2). Marine protist networks included mainly OTUs occurring in six to eight seas and oceans. Terrestrial protist networks included mostly OTUs occurring in two to three forests, and candidate OTUs occurring in a single forest were largely absent from the networks, suggesting that local patterns are disregarded with our approach. The taxonomic coverage of network OTUs remained unchanged in marine datasets compared with candidate OTUs (Supporting Information Figure S3). OTUs of the two clades with the lowest abundance in the terrestrial dataset, Dinophyta and Haptophyta, were not included in the networks, nor were Chlorophyta OTUs in the co-occurrence network and MAST (marine Stramenopiles, polyphiletic basal clade; Massana et al., 2014) OTUs in the co-exclusion network. Phylogenetic closeness (close, intermediate or distant relationship) was determined by the number of shared taxonomic ranks between OTU pairs and showed variations between the marine and terrestrial datasets (Supporting Information Figure S4).

3.2 | Phylogenetic signal in co-occurrence networks

Using null model 1, in which phylogenetic relatedness was drawn randomly from the total pool of candidate OTUs, co-occurring OTUs from the marine datasets had positive SESs that were significant and strong at small to intermediate pairwise sequence distances (< 0.27) and phylogenetic distances (< 1.4), and OTUs from the terrestrial dataset had positive SESs that were significant and strong for some low and intermediate pairwise sequence distances (< 0.25) and for the two lowest phylogenetic distances (< 0.2) (Figure 2). Conversely, OTUs from the marine datasets had negative SESs that were significant for intermediate to large pairwise sequence distances (0.28–0.5) and phylogenetic distances (1.6–5.3 and 6–10 for marine surface, 2.8–3.3 and 5.2–10 for marine DCM), and OTUs from the terrestrial dataset had negative SESs for some intermediate pairwise sequence distances (0.27–0.32) and phylogenetic distance classes (1.4–2.1). Interestingly, co-occurrence in marine DCM showed significantly positive SESs for OTU pairs with phylogenetic distances between 4.1 and 4.7. Likewise, co-occurrence in Neotropical soils showed significant positive SESs for OTU pairs with large pairwise sequence and phylogenetic distances.

Using null model 2, in which the edges were randomized among the set of co-occurring OTUs (Supporting Information Figure S5), similar trends were observed, with the transition between positive and negative SESs taking place at lower pairwise sequence and phylogenetic distances in the marine datasets and with half the number of significant SESs in the terrestrial dataset. These results mean that

TABLE 1 Network parameters

Network	Dataset	Number of samples	Number of candidate OTUs[a]	Number of reads	Spearman's ρ threshold	Number of network OTUs[b]	Number of network reads	Number of candidate correlations[c]	Number of significant correlations[d]	Average network degree	Average network path length
Co-occurrence	Marine surface	47	8,274	1.1×10^8	0.58	4,351 (52.6%)	8.2×10^7	3.4×10^7	49,616 (0.14%)	22.8	4.7
	Marine DCM	32	10,760	6.5×10^7	0.68	3,575 (33.2%)	3.6×10^7	5.8×10^7	25,306 (0.04%)	14.2	6.2
	Neotropical soil	114	687	1.8×10^7	0.45	83 (12.1%)	5.2×10^6	2.4×10^5	373 (0.16%)	9.0	2.2
Co-exclusion	Marine surface	47	8,274	1.1×10^8	−0.52	4,265 (51.5%)	8.1×10^7	3.4×10^7	29,873 (0.09%)	14.0	4.0
	Marine DCM	32	10,760	6.5×10^7	−0.64	3,478 (32.3%)	3.5×10^7	5.8×10^7	13,760 (0.02%)	7.9	5.1
	Neotropical soil	114	687	1.8×10^7	−0.24	41 (6%)	4.8×10^6	2.4×10^5	54 (0.02%)	2.6	3.4

DCM = deep chlorophyll maximum; OTU = operational taxonomic unit.

[a]OTUs of the original dataset occurring in ≥30% of marine or ≥10% of terrestrial samples.

[b]The percentage of candidate OTUs is given in parentheses.

[c]Total number of potential edges between network OTUs.

[d]Number of edges in the network; percentage of potential edges in parentheses.
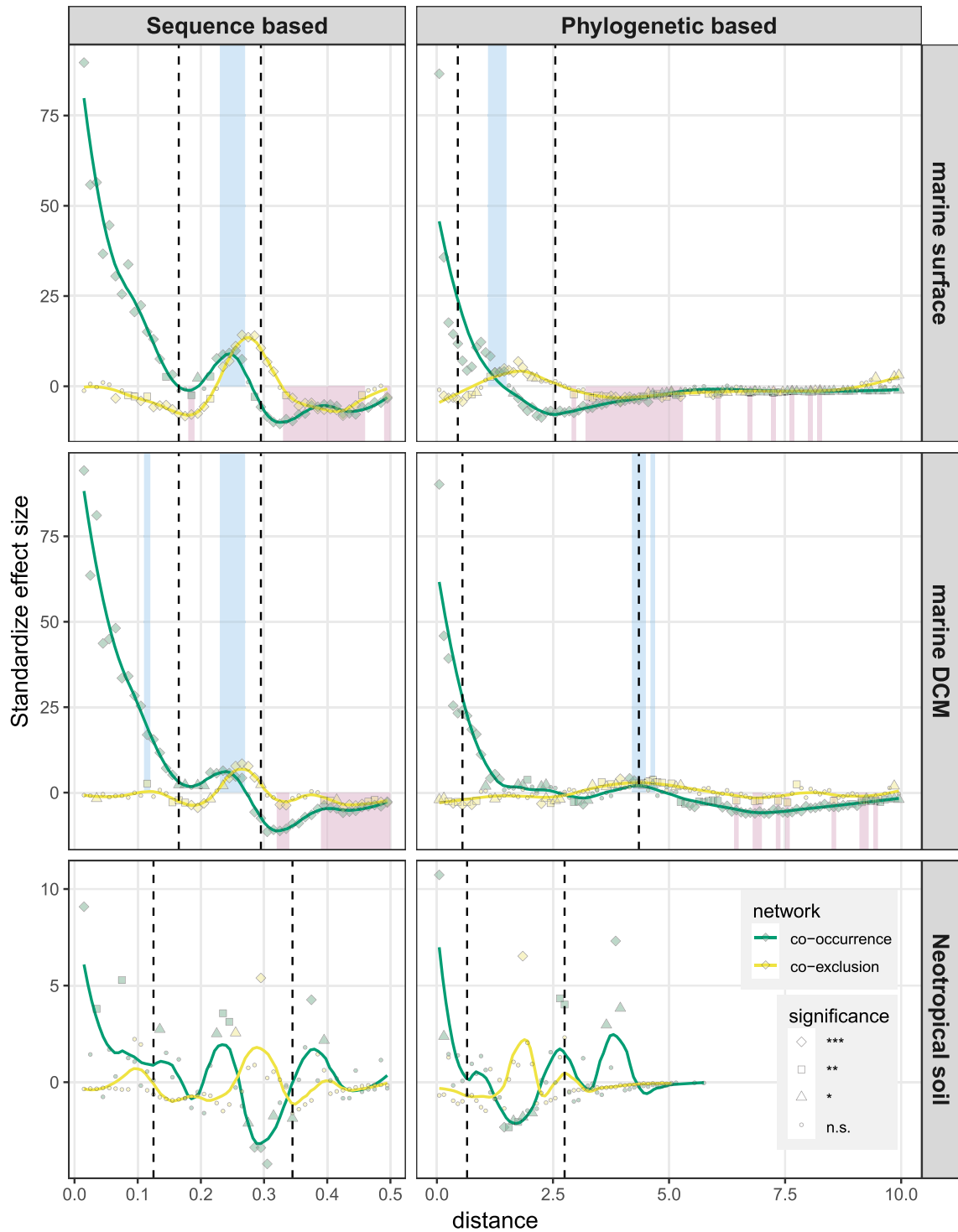
FIGURE 2 Standardized effect sizes (SESs) in co-occurrence and co-exclusion networks compared with null model 1 with shuffled phylogenetic tree tips. SESs were calculated separately for classes of pairwise sequence and phylogenetic distances. Pairs of operational taxonomic units connected by an edge in the observed networks were accounted for in a stepwise manner: from 0 to 0.5 with a step of .01 for pairwise sequence-based distances; and from zero to the maximum phylogenetic distance with a step of .1 for phylogenetically based distances. Two-sided nonparametric *p*-values were computed using 1,000 bootstraps of the null model and are inversely proportional to the number of null models with a higher (for positive SESs) or lower (for negative SESs) amount of co-occurrence or co-exclusion than in the observed network. Values of *p* ≤ .05 were considered significant (**p ≤ .05; **p ≤ .01; ***p ≤ .001; n.s. = not significant). Distance ranges highlighted in blue or red are for distance classes with concurrent excess (significant positive SES) or lack (significant negative SES) of edges in both co-occurrence and co-exclusion networks. Dashed lines delineate distance classes for pairs of operational taxonomic units with close (left), intermediate (middle) and distantrelationships (right) (see Supporting Information Supplementary Material). DCM = deep chlorophyll maximum [Colour figure can be viewed at wileyonlinelibrary.com]

pairs of OTUs with a phylogenetically close or intermediate relationship co-occurred more often than expected by chance in the marine and terrestrial protistan communities, whereas phylogenetically distant OTUs predominantly co-occurred less often than expected by chance. Additionally, for co-occurrences, using either pairwise sequence or phylogenetic distances resulted in similar SES values, with pairwise sequence distance providing a better resolution for taxa with high phylogenetic relatedness.

3.3 | Phylogenetic signal in co-exclusion networks

Using null model 1, co-excluding OTUs from the marine datasets had negative SESs that were significant for low to intermediate pairwise sequence distances (< 0.22) and phylogenetic distances (< 0.9) (Figure 2). Conversely, OTUs from the marine datasets had positive SESs that were significant for intermediate pairwise sequence distances (surface: 0.23–0.32; DCM: 0.23–0.29) and intermediate or large phylogenetic distances (surface: 1.1–2.2; DCM: 3.3–5.3). In the terrestrial dataset, however, no significant SESs were observed except for two pairwise distance classes (between 0.25 and 2.6 and between 2.9 and 0.3) and one phylogenetic distance class (between 1.8 and 1.9) with significant positive SESs.

Using null model 2, similar trends were observed except for the marine DCM, in which significant positive SESs were observed systematically for the majority of the large phylogenetic but not pairwise genetic distance classes (Supporting Information Figure S5). These results mean that pairs of OTUs with a phylogenetic close or intermediate relationship co-excluded less often than expected by chance, whereas phylogenetically distant OTUs co-excluded more often than expected by chance, in the marine protistan communities. In the terrestrial protistan communities, however, phylogenetic relatedness was almost independent of co-exclusion. Additionally, for co-exclusions, as in the co-occurrences, using either pairwise sequence distances or phylogenetic distances in these comparisons resulted in similar SES values, with the same difference in resolution at low phylogenetic distance as observed for co-occurrences and a better detection of significant negative SESs in marine DCM water with phylogenetic distances when using null model 2.

3.4 | Concurrent co-occurrence and co-exclusion at intermediate phylogenetic distances

In all datasets and for most distance classes, positive SESs in co-occurrence networks were reflected by negative SESs in co-exclusion networks and vice versa. The negative SESs in co-exclusion networks for phylogenetically close OTUs were comparatively much lower or non-significant than the positive SESs in the co-occurrence networks. These patterns were confirmed by the edge sampling along distance classes (Supporting Information Figure S6), with co-occurrence networks sampling most of candidate edges in low pairwise sequence and phylogenetic distance values, whereas

co-exclusion networks lacked edges in those low distance values. This implied higher sampling of edges between OTUs from the same genera in the marine datasets or from the same species in the terrestrial dataset for co-occurrence networks (Supporting Information Figure S7). This also resulted in a significantly higher number of shared taxonomic ranks between closely related co-occurring OTUs in marine datasets compared with candidate OTUs pairs for the same distance classes (Mann–Whitney U-test, p < .05; Figure 3; Supporting Information Figure S4).

Null model 1 revealed distance classes for which there were, at the same time, significant positive or negative SESs in both co-occurrence and co-exclusion networks (Figure 2; Supporting Information Figure S5, shaded areas). In particular, positive co-occurrence and co-exclusion SESs were observed at intermediate pairwise sequence (0.24–0.27) and phylogenetic (1.3–1.7) distance classes in the marine surface dataset. In these ranges of distance classes, > 80% of the co-occurrences and co-exclusions were between taxa of different supergroups, with the distribution of edges among shared taxonomic levels not differing significantly from the candidate edges in these same ranges (Figure 3; Supporting Information Figure S4).

A closer look at the taxonomic groups connected by co-occurrences and co-exclusions revealed important shifts in the proportion of edges compared with all candidate edges (Supporting Information Figure S8). Ciliophora were under-represented in both of these co-occurrence and co-exclusion sub-networks compared with all candidate edges, as were Apicomplexa, Bacillariophyta (diatoms), Dinophyta and Radiolaria in the co-occurrence sub-networks, whereas there were increases for almost all other pairs of clades in both sub-networks, in particular for Haptophyta in the co-occurrence sub-network and for Bacillariophyta versus Dinophyta and Haptophyta in the co-exclusion sub-network (Figure 4). Interestingly, there were simultaneous excesses of intra-clade co-occurrences and co-exclusions for Haptophyta, MAST and Telonemia and a simultaneous lack of intraclass co-occurrences and co-exclusions for Ciliophora, Dinophyta and Radiolaria. The amount of change was particularly important when compared with the same sub-networks in the 0.24–027 pairwise sequence distance range of the marine DCM dataset (Supporting Information Figure S9). Edges involved fewer pairs of clades, and the lowest range of fold changes showed a much less divergent sampling of all potential edges than in the marine surface dataset, meaning that no inversion zone was visible for the marine DCM dataset.

4 | DISCUSSION

Network analyses have been widely used in recent environmental sequencing studies to make sense of microbial distributions (Karimi et al., 2019; Milici et al., 2016; Steele et al., 2011). The inclusion of phylogenetic relatedness and null models in the network analyses permits the testing of additional hypotheses on microbial assembly. Here, we assessed the non-random phylogenetic relatedness
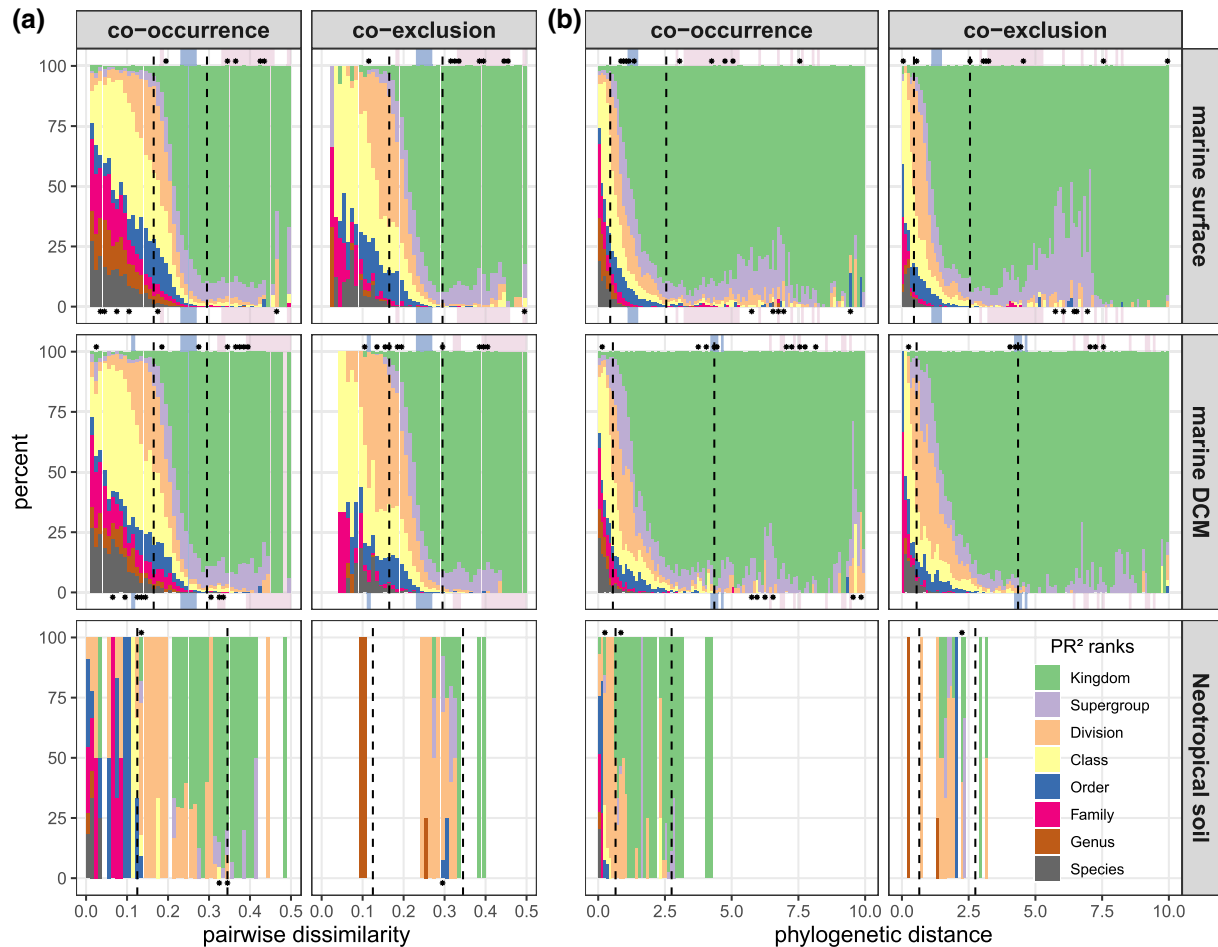
FIGURE 3 Distribution of taxonomic relationships between network connected operational taxonomic units for (a) each pairwise sequence distance class and (b) phylogenetic distance classes. Blue and red shaded areas in the background are the distance classes with simultaneous positive and negative standardized effect sizes (SESs) in both co-occurrence and co-exclusion networks using null model 1, as in Figure 2. Asterisks at the bottom of bars indicate classes with significantly deeper (toward species-level) distribution of taxonomic ranks compared with all candidate edges (Supporting Information Figure S4), whereas asterisks at the top of bars indicate classes with significantly higher (toward domain-level) distribution of taxonomic ranks (Mann–Whitney U-test, $p < .05$). DCM = deep chlorophyll maximum [Colour figure can be viewed at wileyonlinelibrary.com]

of co-occurring and co-excluding OTUs in two of the largest environmental sequencing datasets of marine and terrestrial protists, in order to identify stochastic or deterministic processes responsible for protist assembly. Phylogenetically close OTUs co-occurred more often than expected by chance, and co-occurring OTUs were phylogenetically closer than expected by chance in both environments. The opposite trend was observed for OTUs with intermediate phylogenetic distances, which co-occurred less often than expected by chance, but with much lower intensity. Under the assumption of phylogenetic niche conservatism, these co-occurrence results tend to support the preponderant effect of environmental filtering, which drives protists sharing ecological niche spaces together and separates protists without much overlap in their niche spaces. Recent studies on soil bacteria and peatland testate amoebae also identified environmental filtering as a factor driving closely related species to occur together (Goberna et al., 2019; Singer et al., 2018; Zhang et al., 2018). Dispersal limitation of recently diverging taxa might also play an important role in this pattern, as demonstrated

for the dominant terrestrial protist clades (Lentendu et al., 2018) or in marine ciliates (Azovsky et al., 2020).

Phylogenetically close OTUs were found to co-exclude less often than expected by chance, whereas OTUs with intermediate phylogenetic distances co-excluded more often than expected by chance in the marine environments, almost mirroring the co-occurrence patterns. There was, however, no clear limit between close and intermediate phylogenetic distances, meaning that some distance classes displayed a significant excess of both co-occurrences and co-exclusions at intermediate phylogenetic distances in the marine datasets. In the terrestrial environment, however, co-exclusions were almost independent of phylogenetic relatedness. Under the assumption of phylogenetic niche conservatism, these co-exclusion patterns also reflect the effect of environmental filtering in both marine surface and DCM waters, whereas co-exclusion of protists in Neotropical soils seems to be the result of stochastic processes. One explanation for this discrepancy would be the relatively higher level of homogenization in the marine waters, which allows
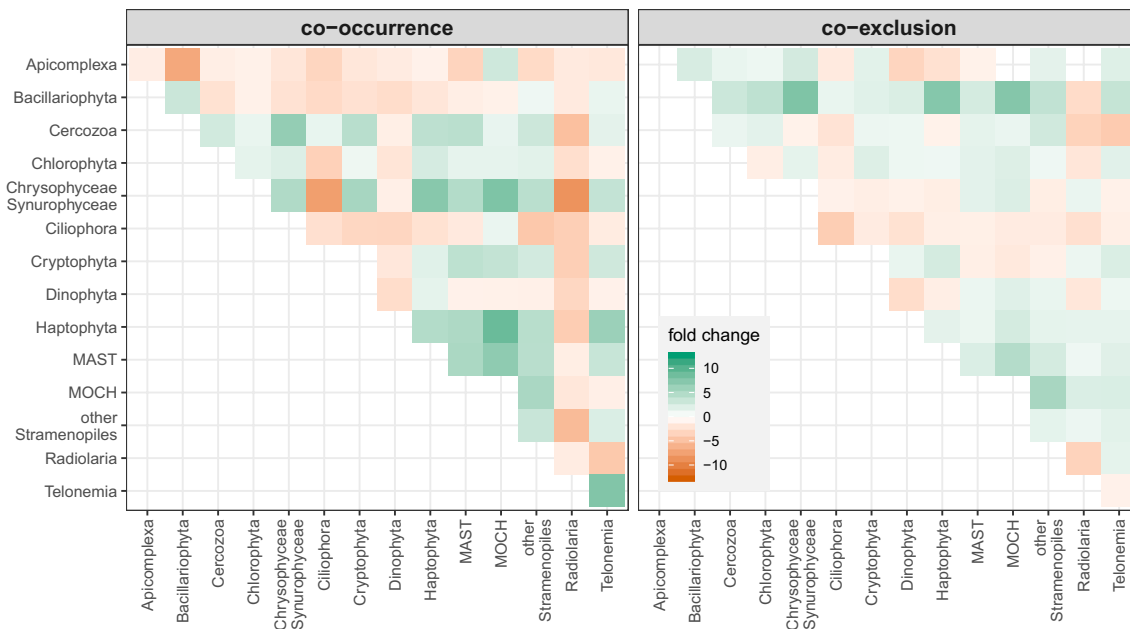
FIGURE 4 Fold changes in the proportion of edges connecting the main clades in the marine surface dataset compared with all candidate edges in the pairwise sequence distance range of 0.24–0.27 (i.e., the largest range of distance with concurrent positive standardized effect sizes in co-occurrence and co-exclusion networks when using null model 1). MAST = marine Stramenopiles; MOCH = marine Ochrophyta [Colour figure can be viewed at wileyonlinelibrary.com]

protists to reach a suitable habitat easily, whereas the larger number of soil protist microhabitats (Adl & Gupta, 2006) and the high local diversity in the Neotropics (Mahé et al., 2017) would blur the impact of environmental filtering and limit potential competitors from coming into contact. Indeed, high local heterogeneity has been shown to promote the coexistence of competing species (Snyder & Chesson, 2004). The difference in spatial scales covered by the global marine datasets and the regional terrestrial dataset could also explain this discrepancy. Dispersal limitation should, in theory, have less impact at the regional scale covered by the terrestrial dataset, but the ease of dispersal in marine waters also reduces the strength of this process even at the global scale. It was previously shown for marine ciliates that competitive exclusion occurred only at a small spatial scale, whereas here we could detect its imprints at the global scale (Azovsky, 1996).

Concurrent excess of positive SESs for co-occurrences and co-exclusions among Haptophyta and Telonemia OTUs with intermediate phylogenetic relatedness could reflect simultaneous effects of environmental filtering and competitive exclusion. This pattern has already been used as explanation for the "paradox of the plankton", which supports the co-occurrence of functionally similar plankton (Huisman & Weissing, 1999; Hutchinson, 1961). Phylogenetically related plankton did co-occur, but they simultaneously co-excluded in marine surface and DCM waters. Biological interactions might also drive this excess of co-occurrences at intermediate phylogenetic distances, because multiple symbiotic associations are extremely common in the marine ecosystem (Bjorbækmo et al., 2020). The dominance of phototrophic taxa in the marine datasets, compared with the dominance of animal pathogen and heterotrophic protists

in the terrestrial dataset, might also explain the presence or absence of the simultaneous excess of co-occurrence and co-exclusion. Biological interactions have been used previously to support patterns of co-occurrence (Goberna et al., 2019). However, we would not risk linking statistically determined co-occurrences to real biological interactions because this has not been validated by live observations in the environments (Lima-Mendez et al., 2015). Other large-scale processes affect the assembly patterns of marine protists, such as the mean annual temperature, responsible for the latitudinal diversity gradient, or the sunlight exposure and currents, responsible for the depth stratification in the water column (Giner et al., 2020; Ibarbalz et al., 2019). Geographical structures, natural fluctuations and the absence of an equilibrium state in marine plankton communities are not enough to avoid co-exclusion inside a clade, as observed here.

There are three novel aspects to the present study. The first novel aspect was the use of null models to test the significance of phylogenetic relatedness structures in co-occurrence and co-exclusion networks. Testing the divergence from the null expectation is needed to distinguish between stochastic and deterministic origins of the observed patterns (Chase & Myers, 2011; Hardy, 2008). Only the relationships between co-occurring/co-excluding protistan OTUs and their putative function or the change in network topology among habitats have been tested previously in marine (Guidi et al., 2016; Lima-Mendez et al., 2015; Milici et al., 2016; Steele et al., 2011), freshwater (Debroas et al., 2017; Posch et al., 2015) and terrestrial environments (Lentendu et al., 2014; Ma et al., 2016; Xiong et al., 2018). In a network-based study on the human microbiome combining analyses of phylogenetic relatedness and co-occurrence/co-exclusion

networks, it was shown that co-occurrences between human bacterial OTUs were distributed uniformly among phylogenetic distances, whereas co-exclusions were mainly among phylogenetically distant OTUs (Faust et al., 2012). In contrast, pairs of co-occurring soil bacteria associated with environmental filtering were found to be more closely related than co-occurring soil bacteria associated with putative biological interactions (Goberna et al., 2019). The lack of a null model and/or statistical test on these observations, however, did not allow those authors to determine whether stochastic or deterministic processes were responsible of the observed patterns.

The second novel aspect is that we showed that both phylogenetic distance and pairwise sequence distance can be used as measures of phylogenetic relatedness when applied to the analysis of protistan community assembly patterns. Pairwise sequence distances provided higher resolution for closely related taxa, whereas phylogenetic distances provided better resolution for distantly related taxa. Previous studies on protists used the phylogenetic relatedness of protists to assess phylogenetic diversity-based macroecological and biogeographical patterns (Bates et al., 2013; Lentendu et al., 2018; Singer et al., 2018), and the pairwise sequence distances were used only during the bioinformatic procedure for sequence clustering or sequence similarity networks (Forster et al., 2019; Mahé et al., 2015).

The third novel aspect was the decomposition of the co-occurrence and co-exclusion signals along phylogenetic distance classes. Using a traditional index of phylogenetic divergence (e.g., net relatedness index), only one type of divergence could be assessed per sample or pair of samples (i.e., either clustering or overdispersion). For example, pairs of bacteria linked in global gut microbiome co-occurrence networks were found to have significantly higher phylogenetic relatedness than in randomized networks overall (Tackmann et al., 2019), whereas a size effect was not quantified for distinct distance classes. Using the co-occurrence and co-exclusion patterns for all samples, here we investigated the multiple signals held by communities over increasing phylogenetic distances for the whole regions analysed. We could show that excesses of both co-occurrence and co-exclusion take place at the same time for only a small range of intermediate phylogenetic distances. We advocate here for a finer analysis of association networks along with phylogenetic relatedness to disclose otherwise hidden patterns.

The main assumption in the present study was that there is phylogenetic niche conservatism between the OTUs (Wiens & Donoghue, 2004). This assumption allowed us to infer that phylogenetically close OTUs share more niche space than phylogenetically distant OTUs. This assumption allows us to interpret the significant excess in co-occurrence among phylogenetically close OTUs as a signal of environmental filtering, and the absence of a significant effect size in co-exclusion among phylogenetically close OTUs as a signal for lack of environmental filtering and competitive exclusion. However, the assumption that evolutionarily close OTUs share the same niche might not be true, and it could be misleading to deduce pattern from process (Gerhold et al., 2015). In such a large dataset, there is a multitude of niche evolution scenarios that lead to the current distribution

of protists in marine waters and Neotropical soils, and the apparent environmental filtering deduced here from the co-occurrence patterns could hide other processes at play that are not necessarily linked to phylogenetic niche conservatism. A modelling approach could also help to test for the reality of phylogenetic niche conservatism by protists (Münkemüller et al., 2015) but remains inapplicable to large datasets, such as analysed here, for which a large proportion of organisms are unknown (de Vargas et al., 2015; Mahé et al., 2017). Considering that current knowledge on traits and function is not sufficient to determine the functional niche of most protists (Ramond et al., 2019), relating phylogeny to assembly patterns with the phylogenetic niche conservatism assumption is the most precise approach we can currently apply to find clues about large-scale and whole-community processes at play in protist community assembly.

By demonstrating the strong phylogenetic signals in co-occurrence and co-exclusion patterns of protists, we show that global and regional assembly mechanisms are related directly to phylogenetic relatedness and are dominated by environmental filtering. We could not conclude that the simultaneous excess of co-occurrence and co-exclusion of phylogenetically related OTUs in the SES inversion zone of the marine surface communities is the result of intraclade competitive exclusion, but we could suspect it. Indeed, multiple other processes could lead to such a pattern, such as facilitation of phylogenetically distant species (Cahill et al., 2008; Gerhold et al., 2015; Kraft et al., 2007). The discrepancy in co-exclusion between marine and terrestrial protists highlights the difference in mechanisms involved in community assembly between these two environments. The novel network phylogeny approach presented in this study has the potential to unravel phylogenetically driven assembly patterns in large-scale datasets for which little is known about the taxonomy and function of the target organisms in other environments.

DATA AVAILABILITY STATEMENT

All scripts and data are available at Dryad (https://doi.org/10.5061/dryad.j6q573ndc). The OTU matrices used in this study and originating from previously published studies are provided in the standard BIOM format (Supporting Information Datasets S1 and S2). The complete bash and R scripts to reproduce the analyses are provided in HTML format (Supporting Information Datasets File S1). The full network calculation procedure is also available as stand-alone software, which can be used on any datasets apart from the one used in this study (https://github.com/lentendu/NetworkNullHPC).

ORCID

Guillaume Lentendu ![ORCID] https://orcid.org/0000-0003-3826-6253
Micah Dunthorn ![ORCID] https://orcid.org/0000-0003-1376-4109

REFERENCES

Adl, M. S., & Gupta, V. S. (2006). Protists in soil ecology and forest nutrient cycling. *Canadian Journal of Forest Research*, *36*, 1805–1817. https://doi.org/10.1139/x06-056

Azovsky, A. I. (1996). The effect of scale on congener coexistence: Can molluscs and polychaetes reconcile beetles to ciliates? *Oikos*, *77*, 117–126. https://doi.org/10.2307/3545591

Azovsky, A. I., Chertoprud, E. S., Garlitska, L. A., Mazei, Y. A., & Tikhonenkov, D. V. (2020). Does size really matter in biogeography? Patterns and drivers of global distribution of marine micro- and meiofauna. *Journal of Biogeography*, *47*, 1180–1192. https://doi.org/10.1111/jbi.13771

Bates, S. T., Clemente, J. C., Flores, G. E., Walters, W. A., Parfrey, L. W., Knight, R., & Fierer, N. (2013). Global biogeography of highly diverse protistan communities in soil. *The ISME Journal*, *7*, 652–659. https://doi.org/10.1038/ismej.2012.147

Bjorbækmo, M. F. M., Evenstad, A., Røsæg, L. L., Krabberød, A. K., & Logares, R. (2020). The planktonic protist interactome: where do we stand after a century of research? *The ISME Journal*, *14*(2), 544–559. http://dx.doi.org/10.1038/s41396-019-0542-5

Boenigk, J., Wodniok, S., Bock, C., Beisser, D., Hempel, C., Grossmann, L., Lange, A., & Jensen, M. (2018). Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *Metabarcoding and Metagenomics*, *2*, e21519. https://doi.org/10.3897/mbmg.2.21519

Burns, J. H., & Strauss, S. Y. (2011). More closely related species are more ecologically similar in an experimental test. *Proceedings of the National Academy of Sciences USA*, *108*, 5302–5307. https://doi.org/10.1073/pnas.1013003108

Cahill, J. F., Kembel, S. W., Lamb, E. G., & Keddy, P. A. (2008). Does phylogenetic relatedness influence the strength of competition among vascular plants? *Perspectives in Plant Ecology, Evolution and Systematics*, *10*, 41–50. https://doi.org/10.1016/j.ppees.2007.10.001

Chase, J. M., & Myers, J. A. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 2351–2363. https://doi.org/10.1098/rstb.2011.0063

Chow, C.-E. T., Kim, D. Y., Sachdeva, R., Caron, D. A., & Fuhrman, J. A. (2014). Top-down controls on bacterial community structure: Microbial network analysis of bacteria, T4-like viruses and protists. *The ISME Journal*, *8*, 816–829. https://doi.org/10.1038/ismej.2013.199

Chung, F., & Lu, L. (2002). Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, *6*, 125–145. https://doi.org/10.1007/PL00012580

Connor, N., Barberán, A., & Clauset, A. (2017). Using null models to infer microbial co-occurrence networks. *PLoS One*, *12*, e0176751. https://doi.org/10.1371/journal.pone.0176751

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray.

de Menezes, A. B., Prendergast-Miller, M. T., Richardson, A. E., Toscas, P., Farrell, M., Macdonald, L. M., Baker, G., Wark, T., & Thrall, P. H. (2015). Network analysis reveals that bacteria and fungi form modules that correlate independently with soil parameters. *Environmental Microbiology*, *17*, 2677–2689. https://doi.org/10.1111/1462-2920.12559

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Bescot, N. L., Probert, I., Carmichael, M., Poulain, J.,

Romac, S., Colin, S., Aury, J.-M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., … Velayoudon, D. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605. https://doi.org/10.1126/science.1261605

Debroas, D., Domaizon, I., Humbert, J.-F., Jardillier, L., Lepère, C., Oudart, A., & Taïb, N. (2017). Overview of freshwater microbial eukaryotes diversity: A first analysis of publicly available metabarcoding data. *FEMS Microbiology Ecology*, *93*(4). https://doi.org/10.1093/femsec/fix023

del Campo, J., Mallo, D., Massana, R., de Vargas, C., Richards, T. A., & Ruiz-Trillo, I. (2015). Diversity and distribution of unicellular opisthokonts along the European coast analysed using high-throughput sequencing. *Environmental Microbiology*, *17*, 3195–3207. https://doi.org/10.1111/1462-2920.12759

Den Boer, P. J. (1980). Exclusion or coexistence and the taxonomic or ecological relationship between species. *Netherlands Journal of Zoology*, *30*, 278–306. https://doi.org/10.1163/002829679X00421

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., & Huttenhower, C. (2012). Microbial co-occurrence relationships in the human microbiome. *PLoS Computational Biology*, *8*, e1002606. https://doi.org/10.1371/journal.pcbi.1002606

Forster, D., Lentendu, G., Filker, S., Dubois, E., Wilding, T. A., & Stoeck, T. (2019). Improving eDNA-based protist diversity assessments using networks of amplicon sequence variants. *Environmental Microbiology*, *21*, 4109–4124. https://doi.org/10.1111/1462-2920.14764

Gause, G. F. (1934). *The struggle for existence*. Williams and Wilkins.

Gerhold, P., Cahill, J. F., Winter, M., Bartish, I. V., & Prinzing, A. (2015). Phylogenetic patterns are not proxies of community assembly mechanisms (they are far better). *Functional Ecology*, *29*, 600–614. https://doi.org/10.1111/1365-2435.12425

Giner, C. R., Pernice, M. C., Balagué, V., Duarte, C. M., Gasol, J. M., Logares, R., & Massana, R. (2020). Marked changes in diversity and relative activity of picoeukaryotes with depth in the world ocean. *The ISME Journal*, *14*, 437–449. https://doi.org/10.1038/s41396-019-0506-9

Goberna, M., Montesinos-Navarro, A., Valiente-Banuet, A., Colin, Y., Gómez-Fernández, A., Donat, S., Navarro-Cano, J. A., & Verdú, M. (2019). Incorporating phylogenetic metrics to microbial co-occurrence networks based on amplicon sequences to discern community assembly processes. *Molecular Ecology Resources*, *19*, 1552–1564. https://doi.org/10.1111/1755-0998.13079

Gotelli, N. J., & McCabe, D. J. (2002). Species co-occurrence: A meta-analysis of J. M. Diamond's assembly rules model. *Ecology*, *83*, 2091–2096. https://doi.org/10.1890/0012-9658(2002)083[2091:SCOAMA]2.0.CO;2

Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y., Audic, S., Berline, L., Brum, J. R., Coelho, L. P., Espinoza, J. C. I., Malviya, S., Sunagawa, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Poulain, J., Searson, S., … Gorsky, G. (2016). Plankton networks driving carbon export in the oligotrophic ocean. *Nature*, *532*(7600), 465–470. https://doi.org/10.1038/nature16942

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., … Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*, D597–D604. https://doi.org/10.1093/nar/gks1160

Hardy, O. J. (2008). Testing the spatial phylogenetic structure of local communities: Statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology*, *96*, 914–926. https://doi.org/10.1111/j.1365-2745.2008.01421.x

Horner-Devine, M. C., & Bohannan, B. J. M. (2006). Phylogenetic clustering and overdispersion in bacterial communities. *Ecology*, *87*, S100–S108.

https://doi.org/10.1890/0012-9658(2006)87[100:PCAOI
B]2.0.CO;2

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton University Press.

Huisman, J., & Weissing, F. J. (1999). Biodiversity of plankton by species oscillations and chaos. *Nature*, 402(6760), 407–410. https://doi.org/10.1038/46540

Humboldt, A. V., & Bonpland, A. (1805). *Essai sur la géographie des plantes.* Levrault, Schoell et compagnie.

Hutchinson, G. E. (1961). The paradox of the plankton. *The American Naturalist*, 95, 137–145. https://doi.org/10.1086/282171

Ibarbalz, F. M., Henry, N., Brandão, M. C., Martini, S., Busseni, G., Byrne, H., Coelho, L. P., Endo, H., Gasol, J. M., Gregory, A. C., Mahé, F., Rigonato, J., Royo-Llonch, M., Salazar, G., Sanz-Sáez, I., Scalco, E., Soviadan, D., Zayed, A. A., Zingone, A., … Wincker, P. (2019). Global trends in marine plankton diversity across kingdoms of life. *Cell*, 179, 1084–1097.e21. https://doi.org/10.1016/j.cell.2019.10.008

Karimi, B., Dequiedt, S., Terrat, S., Jolivet, C., Arrouays, D., Wincker, P., Cruaud, C., Bispo, A., Prévost-Bouré, N. C., & Ranjard, L. (2019). Biogeography of soil bacterial networks along a gradient of cropping intensity. *Scientific Reports*, 9, 3812. https://doi.org/10.1038/s41598-019-40422-y

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. https://doi.org/10.1093/molbev/mst010

Kembel, S. W., & Hubbell, S. P. (2006). The phylogenetic structure of a Neotropical forest tree community. *Ecology*, 87, S86–S99. https://doi.org/10.1890/0012-9658(2006)87[86:TPSOAN]2.0.CO;2

Khomich, M., Kauserud, H., Logares, R., Rasconi, S., & Andersen, T. (2017). Planktonic protistan communities in lakes along a large-scale environmental gradient. *FEMS Microbiology Ecology*, 93(4). https://doi.org/10.1093/femsec/fiw231

Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, 29, 592–599. https://doi.org/10.1111/1365-2435.12345

Kraft, N. J. B., Cornwell, W. K., Webb, C. O., & Ackerly, D. D. (2007). Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *The American Naturalist*, 170, 271–283. https://doi.org/10.1086/519400

Lamb, E. G., & Cahill, J. F. Jr. (2008). When competition does not matter: Grassland diversity and community composition. *The American Naturalist*, 171, 777–787. https://doi.org/10.1086/587528

Lauber, C. L., Strickland, M. S., Bradford, M. A., & Fierer, N. (2008). The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biology and Biochemistry*, 40, 2407–2415. https://doi.org/10.1016/j.soilbio.2008.05.021

Lennon, J. T., Aanderud, Z. T., Lehmkuhl, B. K., & Schoolmaster, D. R. (2012). Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology*, 93, 1867–1879. https://doi.org/10.1890/11-1745.1

Lentendu, G., Mahé, F., Bass, D., Rueckert, S., Stoeck, T., & Dunthorn, M. (2018). Consistent patterns of high alpha and low beta diversity in tropical parasitic and free-living protists. *Molecular Ecology*, 27, 2846–2857. https://doi.org/10.1111/mec.14731

Lentendu, G., Wubet, T., Chatzinotas, A., Wilhelm, C., Buscot, F., & Schlegel, M. (2014). Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: A multiple barcoding approach. *Molecular Ecology*, 23, 3341–3355. https://doi.org/10.1111/mec.12819

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L., Cornejo-Castillo, F. M., … Raes, J. (2015). Determinants of community structure in the global plankton interactome. *Science*, 348(6237), 1262073. https://doi.org/10.1126/science.1262073

Ma, B., Wang, H., Dsouza, M., Lou, J., He, Y., Dai, Z., Brookes, P. C., Xu, J., & Gilbert, J. A. (2016). Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *The ISME Journal*, 10, 1891–1901. https://doi.org/10.1038/ismej.2015.261

Macarthur, R., & Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist*, 101, 377–385. https://doi.org/10.1086/282505

Mahé, 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., … Dunthorn, M. (2017). Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nature Ecology & Evolution*, 1, 0091. https://doi.org/10.1038/s41559-017-0091

Mahé, F., Rognes, T., Quince, C., de Vargas, C., & Dunthorn, M. (2015). Swarm v2: Highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3, e1420. https://doi.org/10.7717/peerj.1420

Manel, S., Couvreur, T. L. P., Munoz, F., Couteron, P., Hardy, O. J., & Sonké, B. (2014). Characterizing the phylogenetic tree community structure of a protected tropical rain forest area in Cameroon. *PLoS One*, 9, e98920. https://doi.org/10.1371/journal.pone.0098920

Martiny, A. C., Treseder, K., & Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *The ISME Journal*, 7, 830–838. https://doi.org/10.1038/ismej.2012.160

Martiny, J. B. H., Jones, S. E., Lennon, J. T., & Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *Science*, 350, aac9323. https://doi.org/10.1126/science.aac9323

Massana, R., del Campo, J., Sieracki, M. E., Audic, S., & Logares, R. (2014). Exploring the uncultured microeukaryote majority in the oceans: Reevaluation of ribogroups within stramenopiles. *The ISME Journal*, 8, 854–866. https://doi.org/10.1038/ismej.2013.204

Mayfield, M. M., & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, 13, 1085–1093. https://doi.org/10.1111/j.1461-0248.2010.01509.x

Mercier, C., Boyer, F., Bonin, A., & Coissac, É. (2013) *SUMATRA and SUMACLUST: fast and exact comparison and clustering of sequences.* metabarcoding.org/sumatra

Milici, M., Deng, Z.-L., Tomasch, J., Decelle, J., Wos-Oxley, M. L., Wang, H., Jáuregui, R., Plumeier, I., Giebel, H.-A., Badewien, T. H., Wurst, M., Pieper, D. H., Simon, M., & Wagner-Döbler, I. (2016). Co-occurrence analysis of microbial taxa in the Atlantic Ocean reveals high connectivity in the free-living bacterioplankton. *Frontiers in Microbiology*, 7, 649. https://doi.org/10.3389/fmicb.2016.00649

Morriën, E., Hannula, S. E., Snoek, L. B., Helmsing, N. R., Zweers, H., de Hollander, M., Soto, R. L., Bouffaud, M.-L., Buée, M., Dimmers, W., Duyts, H., Geisen, S., Girlanda, M., Griffiths, R. I., Jørgensen, H.-B., Jensen, J., Plassart, P., Redecker, D., Schmelz, R. M., … van der Putten, W. H. (2017). Soil networks become more connected and take up more carbon as nature restoration progresses. *Nature Communications*, 8, 14349. https://doi.org/10.1038/ncomms14349

Müller, J. P., Hauzy, C., & Hulot, F. D. (2012). Ingredients for protist coexistence: Competition, endosymbiosis and a pinch of biochemical interactions. *Journal of Animal Ecology*, 81, 222–232. https://doi.org/10.1111/j.1365-2656.2011.01894.x

Münkemüller, T., Boucher, F. C., Thuiller, W., & Lavergne, S. (2015). Phylogenetic niche conservatism – common pitfalls and ways forward. *Functional Ecology*, 29, 627–639. https://doi.org/10.1111/1365-2435.12388

Olff, H., & Ritchie, M. E. (1998). Effects of herbivores on grassland plant diversity. *Trends in Ecology and Evolution*, 13, 261–265. https://doi.org/10.1016/S0169-5347(98)01364-0

Philippot, L., Andersson, S. G. E., Battin, T. J., Prosser, J. I., Schimel, J. P., Whitman, W. B., & Hallin, S. (2010). The ecological coherence of high bacterial taxonomic ranks. *Nature Reviews Microbiology*, 8, 523–529. https://doi.org/10.1038/nrmicro2367

Posch, T., Eugster, B., Pomati, F., Pernthaler, J., Pitsch, G., & Eckert, E. M. (2015). Network of interactions between ciliates and phytoplankton during spring. *Frontiers in Microbiology*, 6, 1289. https://doi.org/10.3389/fmicb.2015.01289

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ramond, P., Sourisseau, M., Simon, N., Romac, S., Schmitt, S., Rigaut-Jalabert, F., Henry, N., de Vargas, C., & Siano, R. (2019). Coupling between taxonomic and functional diversity in protistan coastal communities. *Environmental Microbiology*, 21, 730–749. https://doi.org/10.1111/1462-2920.14537

Saleem, M., Fetzer, I., Dormann, C. F., Harms, H., & Chatzinotas, A. (2012). Predator richness increases the effect of prey diversity on prey yield. *Nature Communications*, 3, 1305. https://doi.org/10.1038/ncomms2287

Santoferrara, L., Burki, F., Filker, S., Logares, R., Dunthorn, M., & McManus, G. B. (2020). Perspectives from ten years of protist studies by high-throughput metabarcoding. *Journal of Eukaryotic Microbiology*, 67, 612–622. https://doi.org/10.1111/jeu.12813

Singer, D., Kosakyan, A., Seppey, C. V. W., Pillonel, A., Fernández, L. D., Fontaneto, D., Mitchell, E. A. D., & Lara, E. (2018). Environmental filtering and phylogenetic clustering correlate with the distribution patterns of cryptic protist species. *Ecology*, 99, 904–914. https://doi.org/10.1002/ecy.2161

Snyder, R. E., & Chesson, P. (2004). How the spatial scales of dispersal, competition, and environmental heterogeneity interact to affect coexistence. *The American Naturalist*, 164, 633–650. https://doi.org/10.1086/424969

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steele, J. A., Countway, P. D., Xia, L. I., Vigil, P. D., Beman, J. M., Kim, D. Y., Chow, C.-E., Sachdeva, R., Jones, A. C., Schwalbach, M. S., Rose, J. M., Hewson, I., Patel, A., Sun, F., Caron, D. A., & Fuhrman, J. A. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *The ISME Journal*, 5, 1414–1425. https://doi.org/10.1038/ismej.2011.24

Tackmann, J., Matias Rodrigues, J. F., & von Mering, C. (2019). Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *Cell Systems*, 9, 286–296. e8. https://doi.org/10.1016/j.cels.2019.08.002

Tedersoo, L., Bahram, M., Cajthaml, T., Põlme, S., Hiiesalu, I., Anslan, S., Harend, H., Buegger, F., Pritsch, K., Koricheva, J., & Abarenkov, K. (2016). Tree diversity and species identity effects on soil fungi, protists and animals are context dependent. *The ISME Journal*, 10, 346–362. https://doi.org/10.1038/ismej.2015.116

Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., Grenyer, R., Helmus, M. R., Jin, L. S., Mooers, A. O., Pavoine, S., Purschke, O., Redding, D. W., Rosauer, D. F., Winter, M., & Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, 92, 698–715. https://doi.org/10.1111/brv.12252

Vandermeer, J. H. (1972). Niche theory. *Annual Review of Ecology and Systematics*, 3, 107–132. https://doi.org/10.1146/annurev.es.03.110172.000543

Violle, C., Nemergut, D. R., Pu, Z., & Jiang, L. (2011). Phylogenetic limiting similarity and competitive exclusion. *Ecology Letters*, 14, 782–787. https://doi.org/10.1111/j.1461-0248.2011.01644.x

Weißbecker, C., Wubet, T., Lentendu, G., Kühn, P., Scholten, T., Bruelheide, H., & Buscot, F. (2018). Experimental evidence of functional group-dependent effects of tree diversity on soil Fungi in subtropical forests. *Frontiers in Microbiology*, 9, 2312. https://doi.org/10.3389/fmicb.2018.02312

Wetzel, C. E., Bicudo, D. D. C., Ector, L., Lobo, E. A., Soininen, J., Landeiro, V. L., & Bini, L. M. (2012). Distance decay of similarity in Neotropical diatom communities. *PLoS One*, 7, e45071. https://doi.org/10.1371/journal.pone.0045071

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., Damschen, E. I., Davies, T. J., Grytnes, J.-A., Harrison, S. P., Hawkins, B. A., Holt, R. D., McCain, C. M., & Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, 13, 1310–1324. https://doi.org/10.1111/j.1461-0248.2010.01515.x

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. *Trends in Ecology and Evolution*, 19, 639–644. https://doi.org/10.1016/j.tree.2004.09.011

Xiong, W., Jousset, A., Guo, S., Karlsson, I., Zhao, Q., Wu, H., Kowalchuk, G. A., Shen, Q., Li, R., & Geisen, S. (2018). Soil protist communities form a dynamic hub in the soil microbiome. *The ISME Journal*, 12, 634–638. ismej2017171. https://doi.org/10.1038/ismej.2017.171

Zhang, Q., Goberna, M., Liu, Y., Cui, M., Yang, H., Sun, Q., Insam, H., & Zhou, J. (2018). Competition and habitat filtering jointly explain phylogenetic structure of soil bacterial communities across elevational gradients. *Environmental Microbiology*, 20, 2386–2396. https://doi.org/10.1111/1462-2920.14247

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., Martiny, J. B. H., Sogin, M., Boetius, A., & Ramette, A. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One*, 6, e24570. https://doi.org/10.1371/journal.pone.0024570

BIOSKETCHES

Guillaume Lentendu is a scientific collaborator at the University of Neuchâtel. His main research interests are the biogeography and macroecology of terrestrial microbial eukaryotes. He is currently focusing on protist distribution along elevation gradients and how island biogeography theories might apply to these organisms. He has specialized in the development and use of next-generation sequencing molecular and bioinformatic methods to characterize microbial terrestrial biodiversity.

Micah Dunthorn is a professor at the Natural History Museum, University of Oslo. His main research interests are the diversity, evolution, and ecology of protists. He is also interested in developing and using clustering and phylogenetic placement methods to evaluate environmental DNA sequencing data. Although he works with protists in general, his taxonomic focus is on ciliates, apicomplexans, and microsporidians.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.