Received: 6 September 2021 | Revised: 1 October 2021 | Accepted: 11 October 2021

DOI: 10.1111/ele.13904

ECOLOGY LETTERS cnrs WILEY

LETTER

Towards revealing the global diversity and community assembly of soil eukaryotes

Farzad Aslani[1] | **Stefan Geisen**[2] | **Daliang Ning**[3] | **Leho Tedersoo**[4,5] | **Mohammad Bahram**[1,6]

[1]Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia

[2]Laboratory of Nematology, Wageningen University, Wageningen, The Netherlands

[3]Department of Microbiology and Plant Biology, Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma, USA

[4]Mycology and Microbiology Center, University of Tartu, Tartu, Estonia

[5]College of Science, King Saud University, Riyadh, Saudi Arabia

[6]Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden

Correspondence
Farzad Aslani, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia.
Email: farzad.aslani@ut.ee

Funding information
Svenska Forskningsrådet Formas, Grant/Award Number: 2017-05019; Netherlands Organization for for Scientific Research, Grant/Award Number: 016.Veni.181.078; Eesti Teadusagentuur, Grant/Award Number: PUT1317 and 1399

Editor: Franciska de Vries

Abstract

Soil fungi, protists, and animals (i.e., the eukaryome) play a critical role in key ecosystem functions in terrestrial ecosystems. Yet, we lack a holistic understanding of the processes shaping the global distribution of the eukaryome. We conducted a molecular analysis of 193 composite soil samples spanning the world's major biomes. Our analysis showed that the importance of selection processes was higher in the community assemblage of smaller-bodied and wider niche breadth organisms. Soil pH and mean annual precipitation were the primary determinants of the community structure of eukaryotic microbes and animals, respectively. We further found contrasting latitudinal diversity patterns and strengths for soil eukaryotic microbes and animals. Our results point to a potential link between body size and niche breadth of soil eukaryotes and the relative effect of ecological processes and environmental factors in driving their biogeographic patterns.

KEYWORDS
biogeography, functional traits, latitudinal diversity gradient, metabarcoding, soil eukaryotes

INTRODUCTION

Soil microorganisms and microscopic animals play a pivotal role in ecosystem functioning (Delgado-Baquerizo et al., 2016, 2020; Wagg et al., 2014). While prokaryotes are mostly involved in soil-atmosphere gas exchange and rapid turnover of mineral and simple organic compounds, fungi are the key players in plant nutrition, decomposition, and mediating diseases. Soil animals and the somewhat overlooked protists are essential components of the soil microbiome as top-down regulators of the function, evolution, and structure of microbial communities and food webs (Gao et al., 2019; Geisen et al., 2018). To understand ecosystem functioning, the dynamics of soil organic matter, and to predict the effects of global change, it is necessary to identify the factors that shape the distribution and structure of various eukaryotic groups (Crowther et al., 2019).

A central question in biogeography is how diversity and distribution of taxa are related to environmental conditions across spatial gradients (Medini et al., 2005). Due to rapid diversification and ease of dispersal, microorganisms do not necessarily follow the distribution patterns observed in animals and macroorganisms (Allison & Martiny, 2008; Decaëns, 2010; Martiny et al., 2006). The biogeographic patterns of soil biota may be driven by both deterministic (e.g., effects of climate, soil, biotic interactions) and stochastic processes (e.g., drift) (Martiny et al., 2006); however, the relative influence of these processes may depend on organism traits such

as body size and dispersal rate (Farjalla et al., 2012). Such traits, if linked to biogeographic patterns, can shed light on how organisms respond to environmental change (Green et al., 2008). In addition to trait-based approaches, we can infer community assembly processes from phylogenetic data, i.e., phylogenetic distance reflects ecological niche distance (phylogenetic signal; Losos, 2008). Combining phylogeny with ecological information of species may thus contribute to a better understanding of community assembly. Global analysis of different organism groups simultaneously can provide a unique opportunity to address the role of body size, dispersal, and phylogenetic signal on distribution patterns of soil organisms.

The global biogeography of soil microbes and microscopic animals has only recently begun to gain attention (Fierer et al., 2012; Tedersoo et al., 2014; Bahram et al., 2018; Phillips et al., 2019; van den Hoogen et al., 2019; Oliverio et al., 2020). These studies indicate that multiple abiotic and biotic factors jointly determine the structure of various microbial groups, but the major underlying factors differ among bacteria, fungi, protists, and nematodes. While soil pH is of particular importance in determining bacterial diversity (Bahram et al., 2018; Delgado-Baquerizo et al., 2018), fungi respond most strongly to climate (Tedersoo et al., 2014), protists to soil moisture (Oliverio et al., 2020), and nematodes to soil texture (van den Hoogen et al., 2019). In addition, vegetation type determines the abundance and diversity of microbial groups locally and on a regional scale (Bahram et al., 2020; Geisen et al., 2018; Nielsen et al., 2010; Wilschut et al., 2019), but its strong effect on a global scale is known only for fungi so far (Davison et al., 2015; Tedersoo et al., 2014). Despite our accumulated knowledge about biogeographic patterns of soil biota, the underlying mechanisms of the distribution patterns remain little explored (Xu et al., 2020). Here, we simultaneously examined (1) the ecological processes and environmental factors structuring community composition, (2) the cross-continent and cross-biome community structure, and (3) latitudinal diversity gradients for soil fungi, protists, and animals (including Nematoda, Arthropoda, and Annelida) at the global scale. We hypothesized that different body sizes and niche breadths of organism groups lead to contrasting diversity patterns and assemblage mechanisms.

MATERIAL AND METHODS

Sampling and molecular analysis

We used a global set of samples that were collected from plots of homogeneous vegetation, minimally affected by humans, following a standardized sampling and processing scheme (Tedersoo et al., 2014). We selected the 193 plots out of 365 based on geographical evenness (to minimize spatial autocorrelation) and high-quality DNA (Table S1). We note that although our sampling effort covered major biomes across all continents except Antarctica, our sampling focused mainly on forest areas, which cover one-third of the global land area, and there are some areas that remain poorly represented in our dataset (Figure S1). In brief, 40 soil subsamples (5 cm diameter to 5 cm depth) were collected from each study plot (2500 m^2), pooled, air-dried, and homogenized. Altogether, a 2.0 g amount of each of these homogenized composite samples was subjected to DNA extraction using the PowerMax Soil DNA Isolation Mini kit (MoBio) following the manufacturer's instructions (Tedersoo et al., 2014). We used universal eukaryote primers 1389f and 1510r in the polymerase chain reaction (PCR) mix to amplify the V9 region of the 18S rRNA gene (Amaral-Zettler et al., 2009). Forward and reverse primers were indexed with 10-base to 12-base unique multiplex identifiers. The PCR mixture was prepared with 0.3 µl DNA extract, 0.5 µl each of the primers, 5 µl 5xHOT FIREPol Blend Master Mix (Solis Biodyne), and 16 µl double-distilled water. We performed PCR using the following thermocycling conditions: 95°C for 15 min, 30 cycles of 95°C for 30 s, 50°C for 45 s and 72°C for 1 min, with a final extension step at 72°C for 10 min. Amplicon pools were quality-checked and quantified using Bioanalyzer HS DNA Analysis Kit (Agilent) and Qubit 2.0 Fluorometer with dsDNA HS Assay Kit (Thermo Fisher Scientific), respectively, and sequenced on an Illumina HiSeq 2500 platform (2 × 250 paired-end mode) at the Estonian Genome Center (Tartu, Estonia) following Bahram et al. (2018).

Bioinformatics

Quality filtering was performed using the LotuS pipeline (Hildebrand et al., 2014) as outlined in Bahram et al. (2018). Briefly, reads were demultiplexed and quality-filtered based on the following settings: trimming of reads after an accumulated error of 1, rejecting reads of average quality <28 and estimated accumulated error >2.5 (probability ≥0.01). Chimeric reads were removed using both de novo and reference-based chimera checking algorithms and the RDP reference database (http://drive5.com/uchime/rdp_gold.fa) in UCHIME (Edgar, 2011). Out of 7,462,813 reads, 2,089,653 passed quality control. The passed reads were clustered with UPARSE (Edgar, 2013) at 97% sequence similarity. Representative sequences for each non-singleton OTUs were chosen for taxonomic assignment by aligning full-length sequences with lambda (Hauswedell et al., 2014) to the SILVA v.123 database (Pruesse et al., 2007) and using the LotuS least common ancestor (LCA) algorithm (Hildebrand et al., 2014). Based on taxonomic assignments, we selected fungi, protists, and animals for further analyses.

Data analysis

Determination of body size and niche breadth of organism groups

To infer the relative importance of ecological processes on organism groups with various body sizes and niche breadths, we selected the most abundant phyla representing ≥10% of the total fungal, protist and animal reads. The average body size of each phylum was obtained from Briones (2014), Zinger et al. (2019) and Luan et al. (2020). To determine the average niche breadth for each organism group, we calculated the niche breadth for each OTU based on the Levins' index (Levins, 1968) as implemented in *niche.width* function of *spaa* package (Zhang, 2016).

Phylogenetic tree construction

To calculate phylogenetic distance between all pairs of OTUs, we generated sequence alignments of representative sequences of OTUs using mafft (version 7; Katoh & Standley, 2013), followed by masking alignment to minimize alignment ambiguity (Lane, 1991) with default parameters (including maximum relative frequency of gap characters in a column = 1; minimum relative frequency of at least one non-gap character in a column = 0.4). Following this, we built the phylogenetic trees using RAxML (version 8) with a GTRCAT model with 100 bootstrapped replicates (Stamatakis, 2014). Using the generated tree, we computed distances between all pairs of tips of the phylogenetic tree using *distTips* function in the *adephylo* package (Jombart & Dray, 2008).

Null model analysis

To infer the relative effects of ecological processes, we used Community Assembly Mechanisms by Phylogenetic bin-based null model analysis (iCAMP) framework developed by Ning et al. (2020) from a previous framework (Stegen et al., 2013). Individual populations in a community might differently respond to ecological processes (Caruso et al., 2011; Hanson et al., 2012; Ning et al., 2020). Therefore, the iCAMP framework quantifies the relative importance of different ecological processes for each phylogenetic group (bin) rather than only for the entire community, leading to obvious improvement in quantitative performance (Ning et al., 2020). Here we divided OTUs into different groups ('bins') based on their phylogenetic relationships (phylogenetic binning). Then to assess phylogenetic signal, we calculated Pearson correlation between the pairwise phylogenetic distances and niche preference differences for each individual bin with Mantel test (Table S2). Finally, we performed an abundance-based null model analysis based on a phylogenetic dissimilarity metric using beta Net Relatedness Index (βNRI) and taxonomic dissimilarity metric using Bray–Curtis-based Raup–Crick (RC_{bray}) (Stegen et al., 2012, 2013) for each bin. These methods enabled us to evaluate the deviation between the observed phylogenetic/Bray–Curtis dissimilarity and the null-expected phylogenetic/Bray–Curtis dissimilarity. To generate null expectations of community dissimilarities for each sample pair, average phylogenetic and Bray–Curtis dissimilarities of 999 randomly assembled pairs of communities were calculated. The fraction of pairwise comparisons across communities (samples) with $|\beta NRI| > 1.96$ was considered a selection threshold. The RC metric was applied for pairwise comparisons with $|\beta NRI| \leq 1.96$. The fraction of pairwise comparisons with $|RC| > 0.95$ was considered an indicative of dispersal limitation or homogenizing dispersal, whereas with $|RC| \leq 0.95$ was interpreted as the contribution of drift (ecological drift and other processes such as stochastic speciation, weak selection, normal-rate stochastic dispersal). The fractions of ecological processes across all bins were weighted by the relative abundance of each bin and integrated to obtain the relative importance of ecological processes at the whole community level.

The output of null-model-based approaches depends on the sampling effort and species pool setting (Chase & Myers, 2011). Comparing two communities with different regional species pool sizes, the absolute magnitude of the deviation from the null model expectation would be higher (showing stronger deterministic effects) in the community with a larger species pool size. To overcome this weakness, we modified iCAMP framework to test different regional pool settings to count species in each continent sharing the same regional pool (R script is provided in the supplementary files). We set each continent as a regional pool in the null model algorithm. We calculated the relative importance (%) of selection (e.g. heterogeneous selection and homogeneous selection), dispersal (e.g. dispersal limitation and homogenizing dispersal), and drift for each pair of communities (samples) and obtained the mean of percentage of ecological processes for each organism group. Following this, we compared the mean of each individual process among organism groups using a Kruskal-Wallis Test.

Multivariate analysis

The OTU-by-sample matrix was Hellinger-transformed and standardized using the function *decostand* in *vegan* package (Oksanen et al., 2020) in R statistical computing environment (v.3.6.1). Of highly correlated variables ($R > 0.90$), those that explained relatively less community variation were removed to reduce collinearity. To evaluate the extent of spatial autocorrelation, geographical coordinates of plots were transformed into principal

coordinates of neighbor matrices (PCNM) eigenvectors using *vegan* and *packfor* (Dray et al., 2016) packages.

Mantel tests—as implemented in *vegan* package—were conducted to determine the correlation between community structure dissimilarity (Bray-Curtis), environmental dissimilarity (Euclidean) matrices, and geographical distance. We computed the geographic distance between sampling sites using *distm* function of the *geosphere* package (Hijmans, 2019). To disentangle explained (e.g., the unique and shared effects of environmental and spatial vectors) and unexplained variation in phylogenetic dissimilarity matrix, we conducted variation partitioning separately for each organism group using *vegan* package. To test the effects of biome and continent on community structure, we used permutational multivariate analysis of variance (PERMANOVA) with Bray-Curtis dissimilarity and 999 permutations as implemented in *adonis* function in the *vegan* package.

Regression analysis

To assess latitudinal diversity patterns and the relationship between OTU diversity (Shannon index) and environmental gradients, linear and polynomial regressions were chosen based on the adjusted coefficient of determination (R^2_{adj}) and visualized using *ggplot2* package (Wickham, 2016). Shannon diversity index was calculated based on residuals of OTU diversity in relation to the square root of the number of obtained sequences to account for differences in sequencing depth according to Tedersoo et al. (2014). Relative abundances of dominant phyla were compared across biomes using the Kruskal–Wallis test followed by Benjamini–Hochberg's correction for multiple comparisons, as implemented in *dplyr* package (Wickham et al., 2020).

RESULTS

Taxonomic profile of soil eukaryome

Altogether 56.6%, 11.3%, and 17.7% of reads were assigned to fungi, protists, and animals, respectively (Figure S2). Fungal reads were clustered into 2105 OTUs including 1000 Ascomycota (47.5% OTUs; 36.0% fungal reads) and 730 Basidiomycota (34.6% OTUs; 60.0% fungal reads). The 2558 protist OTUs belonged to 7 kingdoms, with SAR supergroups (Stramenopila, 9.91%; Alveolata, 43.99%; Rhizaria, 36.34%) and Amoebozoa (6.24%) accounting for 96.48% of protist reads and 93.57% of OTUs. Animals comprised 1143 OTUs, with Annelida (5.7% OTUs; 17.3% reads), Arthropoda (48.3%; 54.3%), Nematoda (23.2%; 11.5%), and Rotifera (4.0%; 8.7%) collectively accounting for 81.0% of OTUs and 86.0% of reads (Figure S2). The relative abundance of these groups varied among biomes (Table S3). Basidiomycota was significantly more abundant in tropical forests compared with boreal forests ($p < 0.05$). Alveolata was the dominant group in tropical forests, Rhizaria prevailed in Mediterranean biomes, and Stramenopila was relatively more abundant in temperate forests compared with tropical forests ($p < 0.05$). Of animals, Arthropoda and Annelida were relatively most abundant in tropical forests compared with temperate and savanna biomes. By contrast, Nematoda and Rotifera were most abundant in savannas (Table S3).

Relative effects of ecological processes on community assembly

On average, larger organisms had a narrower niche compared with smaller groups, as reflected in the negative correlation between body size and niche breadth ($r = -0.803$, $p < 0.05$). Our analysis showed that drift was the most important ecological process driving the community assembly of all organism groups. Dispersal was the second most important ecological processes for all organism groups, except for the smallest-bodied groups, Rhizaria and Basidiomycota. Drift affected more strongly animal groups (78.0%, 66.2%, and 75.5% for Annelida, Arthropoda, and Nematoda, respectively) compared with protist groups (47.9%, 65.9%, and 59.7% for Alveolata, Stramenopiles, and Rhizaria, respectively) and fungal groups (44.1% and 38.4% for Ascomycota and Basidiomycota, respectively) (Figure 1). This points to a greater role of drift in shaping animal assemblages compared to eukaryotic microbes (Figure 2a). By contrast, the relative importance of selection was higher for the smaller-bodied organism groups such as fungi (25.9%, 36.6% for Ascomycota and Basidiomycota, respectively) and protists (18.1%, 14.1%, and 20.4% for Alveolata, Stramenopiles, and Rhizaria, respectively) compared with animals (8.9%, 6.7% and 3.7% for Annelida, Arthropoda, and Nematoda, respectively). We found that body size was positively related to the proportion of drift ($R^2_{adj} = 0.146$, $p < 0.01$) and negatively related to selection ($R^2_{adj} = 0.195$, $p < 0.01$) in community assembly (Figure 1). By contrast, niche breadth was negatively related to drift ($R^2_{adj} = 0.073$, $p < 0.01$) and positively to selection ($R^2_{adj} = 0.115$, $p < 0.01$) (Figure 1). Variation partitioning analyses supported this finding: more than 50% of the community structure of microbes (52% and 60% for fungi and protists, respectively) was explained by environmental factors and spatial vectors and their shared effects, whereas a large proportion of the community variation of Annelida (79%), Arthropoda (61%), Nematoda (62%), and all animals (69%) remained unexplained (Figure 2b). Similarly, Mantel tests indicated that fungi and protists were more strongly correlated with environmental dissimilarity matrix (Mantel $r = 0.440$ and $r = 0.512$ respectively; $p < 0.001$) compared with Nematoda ($r = 0.275$; $p < 0.001$), Arthropoda ($r = 0.239$;

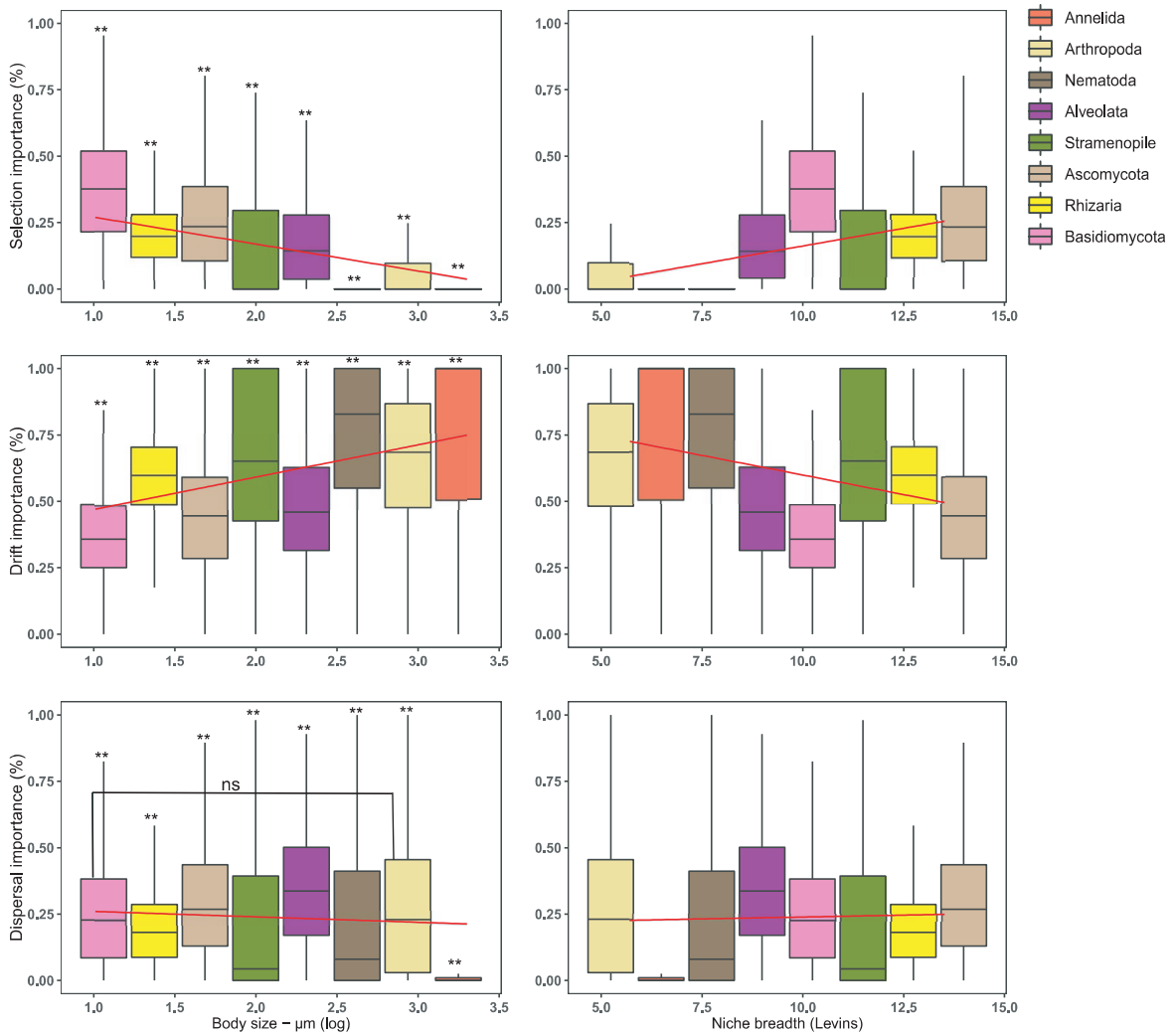

FIGURE 1 Box plots showing the importance of ecological processes in the community assembly of organism groups with various body sizes and niche breadths. Red lines demonstrate a relationship between the importance of ecological processes and body size, as well as niche breadth. "**" indicates significance and "ns" indicates non-significant differences between organism groups ($p < 0.01$). We performed the Kruskal–Wallis test and multiple comparisons using the Wilcoxon test and corrected with the Bonferroni method

$p < 0.001$), Annelida ($r = 0.076$; $p < 0.001$), and the whole animal community ($r = 0.315$; $p < 0.001$) (Figure 2c). Nevertheless, the community structure of microbes and animals showed comparable correlations with geographic distance (Figure 2c, Figure S3).

Environmental determinants of the global eukaryome

Mantel tests revealed that the structure of soil eukaryome is associated with both soil and climate variables. While the structure of fungal and protist communities was strongly related to soil pH, the community structure of animals was mainly related to mean annual precipitation (MAP), soil moisture and fire history (Table 1). More specifically, the community structure of Nematoda was mostly related to MAP and soil pH, whereas that of Arthropoda was mainly related to fire history, MAP, and soil moisture, and Annelida was more influenced by fire history followed by soil moisture (Table 1).

The PERMANOVA analysis revealed cross-biome differences in most organism groups including fungi (biome effect: $R^2_{adj} = 0.109$, $p < 0.01$), protists ($R^2_{adj} = 0.129$, $p < 0.01$), Nematoda ($R^2_{adj} = 0.127$, $p < 0.01$) and Arthropoda ($R^2_{adj} = 0.098$; $p < 0.01$) but not Annelida (Figure 3). The biome effect was relatively stronger than continent effect for all organism groups (continent effect: fungi, $R^2_{adj} = 0.063$; protists, $R^2_{adj} = 0.066$; animals, $R^2_{adj} = 0.069$; $p < 0.01$).

Latitudinal gradient of diversity

Eukaryotic microbes and animals differed in the latitudinal gradient of diversity. Shannon index of fungi and protists showed a hump-shaped relationship with absolute latitude (Figure 4). By contrast, Arthropoda and the total animal diversity showed a U-shaped pattern,

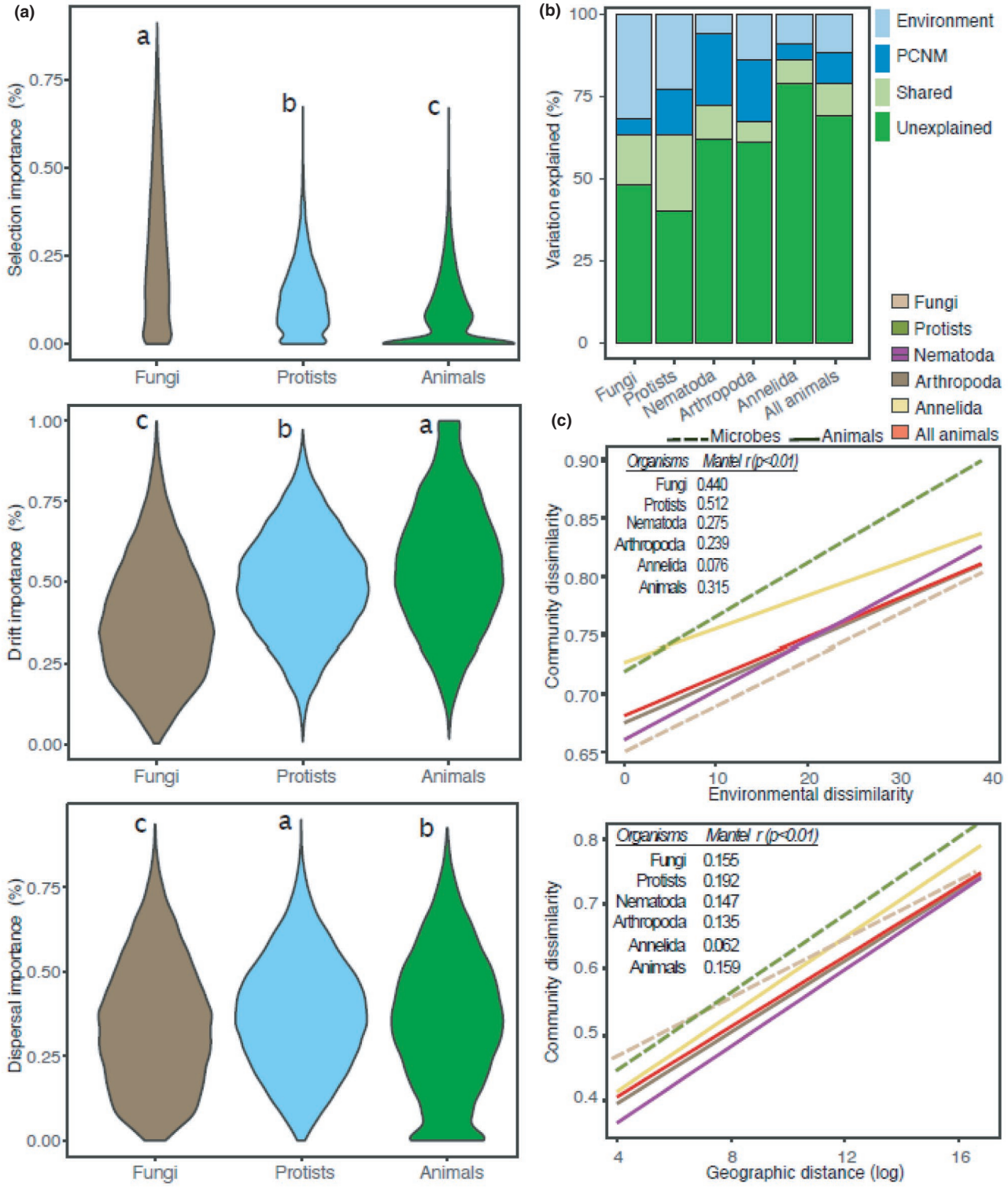

FIGURE 2 (a) Violin plots demonstrate the importance of ecological processes in the assembly of the entire community of fungi, protists, and animals. Organisms with different letters showing significant differences. (b) Barplot demonstrates the explained variation by the effects of environmental factors, PCNM vectors and their shared effects as well as unexplained variations. (c) Relationships between the dissimilarity of eukaryotic communities (Bray-Curtis) and geographic distance (log) as well as environmental distance (Euclidean distance). Environmental variables included soil pH, moisture, carbon, mean annual precipitation, mean annual temperature, fire history, and aridity index. The test statistics of simple Mantel tests are presented for each group

with the highest diversity in tropical forests (Figure 4), whereas the diversity of Nematoda and Annelida decreased linearly towards poles (Figure 4). Our analysis also showed that larger-bodied organisms exhibited a relatively stronger latitudinal gradient of diversity, compared to microbes (Figure 4).

TABLE 1 Correlations between community dissimilarity (Bray-Curtis) and environmental dissimilarity (Euclidean) matrices based on Mantel tests

Variables	Fungi	Protist	Animal	Nematoda	Arthropoda	Annelida
pH	0.412[1,**]	0.407**	0.254**	0.316**	0.169**	0.066*
Moisture	0.217**	0.187**	0.300**	0.220**	0.272**	0.222**
Soil carbon	0.1603**	0.201**	0.194**	0.150**	0.191**	0.139**
MAP	0.269**	0.326**	0.345**	0.319**	0.276**	0.138**
MAT	0.297**	0.378**	0.264**	0.242**	0.217**	0.067**
Aridity index	0.130**	0.167**	0.054ns	0.085*	0.042ns	0.024ns
Fire	0.158**	0.089**	0.298**	0.165**	0.287**	0.324**

Note: Non-significant: ns.

[1]Mantel r.

**$p < 0.01$; *$0.01 < p < 0.05$.

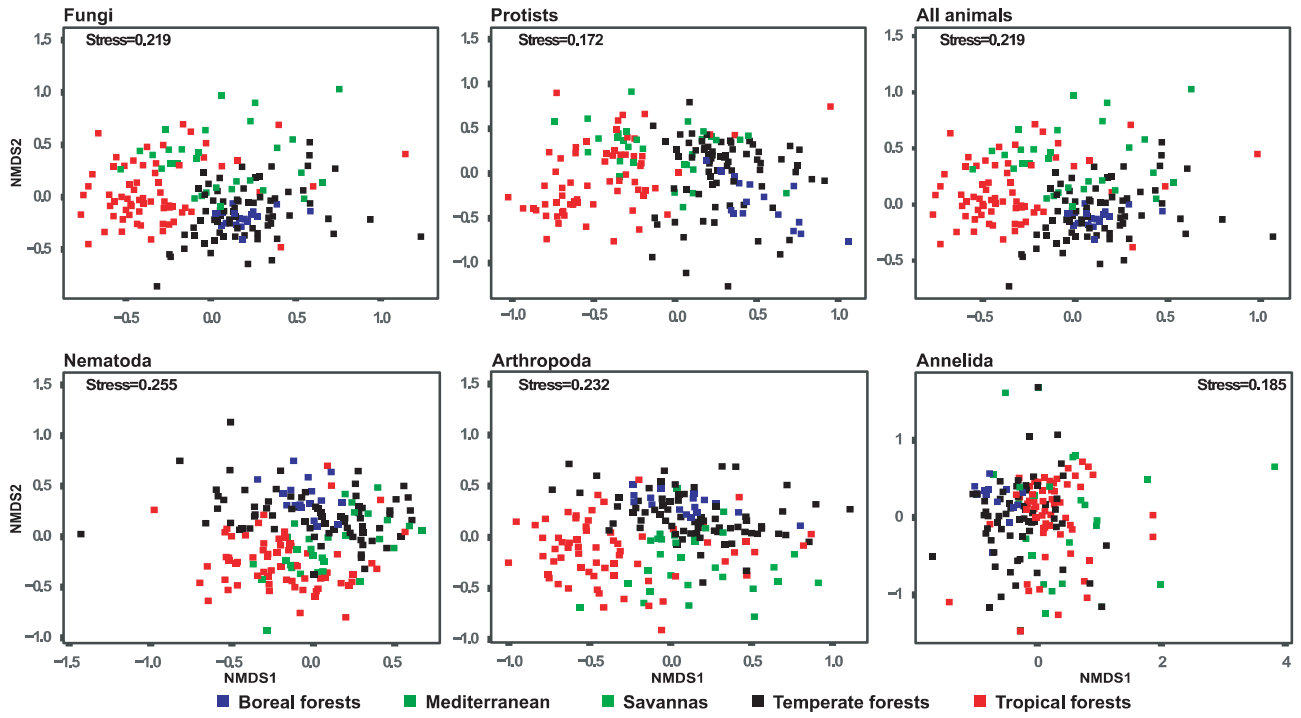

FIGURE 3 Non-metric Multidimensional Scaling (NMDS) plots show the variation in community structure across biomes

Among all tested variables, MAP and soil pH were the strongest predictors of fungal diversity ($R^2_{adj} = 0.058$; $p < 0.001$) and protist diversity ($R^2_{adj} = 0.098$, $p < 0.001$), respectively. By comparison, MAP and mean annual temperature (MAT) were the strongest diversity determinants for Arthropoda (MAP: $R^2_{adj} = 0.056$; MAT: $R^2_{adj} = 0.098$; $p < 0.001$), Annelida (MAP: $R^2_{adj} = 0.174$; MAT $R^2_{adj} = 0.035$; $p < 0.001$), and Nematoda (MAP: $R^2_{adj} = 0.127$; MAT: $R^2_{adj} = 0.016$; $p < 0.001$) (Figure S4).

DISCUSSION

Relative effects of ecological processes

Our data indicate that the relative effects of ecological processes differ among organism groups within the soil eukaryome, which could be partly ascribed to the differences in body size as well as niche breadth. Despite their wider niche breadth and smaller body size, the community structure of fungi and protists was determined more strongly by deterministic processes (heterogeneous and homogeneous selections) than that of animals. This finding suggests that compared to soil animals, microbes with broader niches may be able to adapt to broader ranges of environmental conditions globally. In line with this result, deterministic processes showed relatively stronger effects on the assembly of bacterial communities, with smaller body size and higher dispersal rate, compared to fungal communities (Powell et al., 2015). Higher dispersal rate along with more rapid population growth rates, resulting from a smaller body size, can lead to relatively stronger deterministic processes, through better abilities to arrive at new habitats and faster establishment. In addition, smaller organisms respond more rapidly to

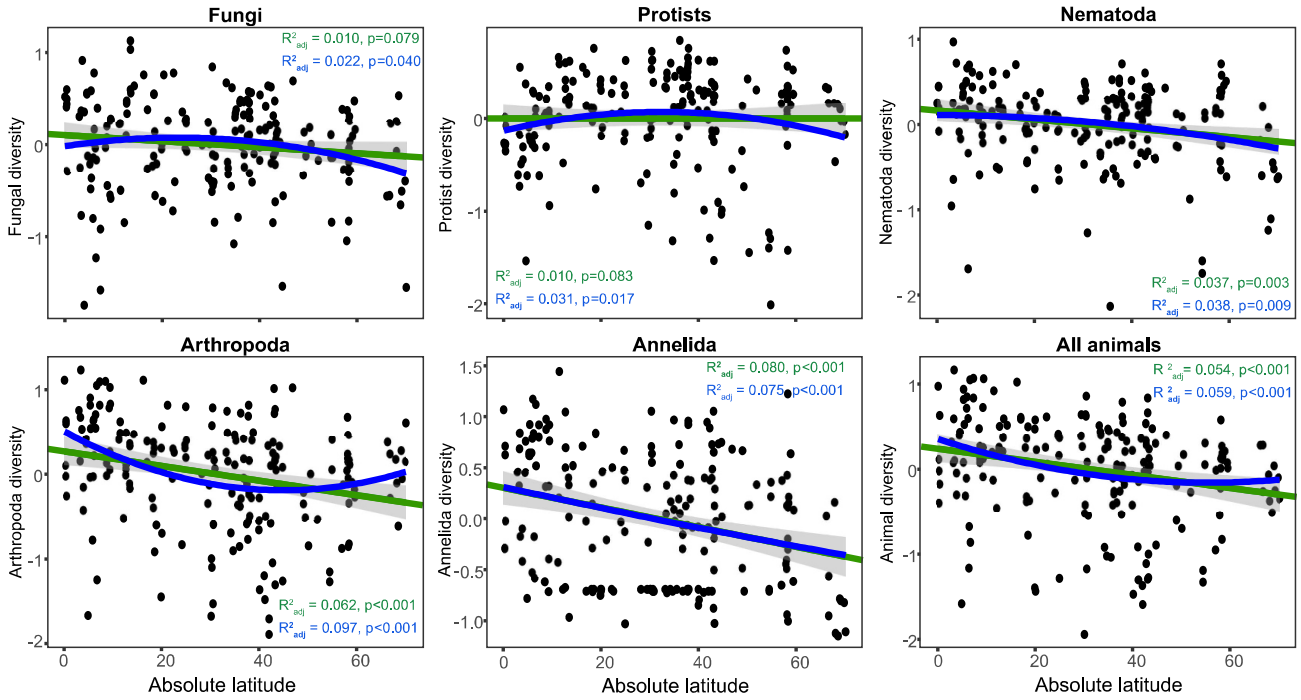

FIGURE 4 Relationships between the diversity of the eukaryotic organisms and latitudinal gradients. First- and second-order polynomial fits are shown in green and blue, respectively. Diversity was measured using the Shannon diversity index

environmental change (Korhonen et al., 2010; Vellend et a., 2014). By contrast, a lower dispersal rate may hamper species to colonize various environmental conditions and thus reduce the effects of environmental selection on community assembly (Leibold et al., 2004). Compared to microbes, animals are known to be less abundant and diverse in soils (Decaëns, 2010), which may contribute to the greater stochasticity in their community structure (Jia et al., 2018) due to their narrower niches (Hanson et al., 2012). Alternatively, larger-bodied, less abundant, and less widespread organisms are probably more prone to extinction (Fodelianakis et al., 2021) and thus show more stochastic distribution patterns compared to smaller-bodied taxa (De Bie et al., 2012; Nemergut et al., 2013; Zinger et al., 2019).

Several studies have shown that larger-bodied organisms with narrower ecological niches are more strongly affected by deterministic processes (Chen et al., 2021; Farjalla et al., 2012; Luan et al., 2020; Soininen et al., 2013). Different ecosystems, geographical scales, and statistical approaches may affect the relative importance of community assemblage processes (Evans et al., 2017; Forbes & Chase, 2002; Hanson et al., 2012; Ladau & Eloe-Fadrosh, 2019; Zhou & Ning, 2017). At the local scale, smaller organisms with higher dispersal rates are commonly ubiquitous and their community assembly is governed by stochastic processes due to small environmental gradients (Bahram et al., 2016). By contrast, broader environmental gradients and more diverse vegetation types could result in stronger environmental filtering of organisms with wider niche breadth. We note that accurate estimates of ecological processes remain a challenge because of the complexity of natural communities and their interactions as well as methodological limitations in inferring these processes. There are several limitations to inferring the relative importance of ecological processes using null-model-based approaches. The result might be varied depending on null model algorithms, similarity metrics for randomization, arbitrary threshold between observed community dissimilarity, and the mean of the null distribution, spatial scale, and regional species pool (Ning et al., 2019). Therefore, the results should be considered as statistical proximate and should be cautiously interpreted on a relative basis (Zhou & Ning, 2017) such as our relative comparison among organism groups.

We also note that since null model-based β-deviation might be influenced by sampling effort (Bennett & Gilbert, 2016; Xing & He, 2021), we performed null model tests with and without rarefaction. Although rarefication led to the overestimation of drift processes, we observed very similar patterns (Figure S5). It is in line with a previous study showing that rarefication, as a random sub-sampling process, added artificial stochasticity to the results of the iCAMP framework, compared to the original communities (see Figure S8a in Ning et al., 2020). Sampling effort might also have an effect on the variations and assembly mechanisms of animal communities, especially for low abundant groups (Jia et al., 2018; Lynch & Neufeld, 2015). In addition, some limitations regarding the used primers and sequencing depth in uncovering certain eukaryotic groups (Tedersoo et al.,

2015), especially the low resolution of 18S region for targeting animal groups (de Groot et al. 2016) may contribute to higher stochasticity in community assembly of animal groups. More sampling sites and sequencing depth together with experimental studies are needed to obtain more confident results for a global-scale assessment.

Further, an adequate phylogenetic signal is necessary for the null model-based approach to infer ecological processes (Stegen et al., 2012, 2013). Within-bin phylogenetic signal test showed that the phylogenetic distance of most (but not all) of the bins of organism groups significantly correlated with the Euclidean distance matrix of at least one environmental factor (Table S2). In the previous study with simulated microbial communities, iCAMP showed robustness to this level of low phylogenetic signal, and the accuracy and precision were still adequate (>0.8) although indeed reduced (Ning et al., 2020).

Environmental determinants of the global eukaryome

Our analysis indicates that biome type, climate, soil factors, and fire history may all affect the eukaryome structure, but their impact differs among major groups of organisms. All studied soil organism groups (except Annelida) exhibited differences in community structure across biomes, but the effect of continents was relatively weaker, supporting the importance of climate and vegetation type in shaping the eukaryome (Bahram et al., 2020; van den Hoogen et al., 2019; Nielsen et al., 2010; Oliverio et al., 2020; Tedersoo et al., 2014; Wilschut et al., 2019).

Our results suggest that organism groups with different body sizes and niche breadths respond differently to environmental variables. The community structure of microbial groups was affected more strongly by soil pH, whereas MAP, soil moisture, and fire history were the main determinants of animal groups (Table 1). Microbial responses to environmental variables have also been shown to depend on gross morphology and microbial domain (Daws et al., 2020). Other studies have reported different environmental variables underlying the distribution of bacteria and fungi (Bahram et al., 2018), as well as between bacteria and protists (Oliverio et al., 2020; Xiong et al., 2021). Thus, it is tempting to speculate that traits such as body size and thereby niche breadth may determine how soil organisms respond to environmental change.

Latitudinal gradient of diversity

Groups of small eukaryotes differed in diversity patterns in relation to latitude (Figure 4), which are partly related to the prevalence of different edaphic and climatic predictors of diversity (Figure S4). Similarly to aboveground macroorganisms (Gaston, 2000) and in line with the known effect of climate on soil animal diversity (Bastida et al., 2020), soil animal diversity increased towards the equator with increasing MAT and MAP. Conversely, the diversity of protists was mainly driven by soil pH (Figure 4; Figure S4), which is only weakly related to latitude. Similarly, a regional-scale study demonstrated that soil properties, particularly soil pH, determined soil microbial diversity but not animal diversity (George et al., 2019). Furthermore, there was a positive relationship between the strength of latitudinal diversity gradient and body size, which corroborates previous meta-analyses on a wide range of organisms (Hillebrand & Azovsky, 2001; Kinlock et al., 2018). Taken together, both the shape and strength of the latitudinal diversity gradient appear to depend on organisms' body size and their associations with environmental variables.

CONCLUSIONS

Our data suggest that drift is a key ecological process in shaping global community assembly of soil eukaryotes, but its relative strength depends on functional traits such as organism's body size and niche breadth. These traits also determine the strength and direction of the association of soil organism groups to environmental effects and latitude. Our findings emphasize the importance of organisms' traits in driving assemblage mechanisms and biogeographic patterns of soil eukaryome. Further work is needed to understand how other traits, especially those related to dispersal and stress resistance, determine responses of these organisms to environmental change.

ACKNOWLEDGMENTS

We thank Oliver Knox and three anonymous reviewers for their constructive comments on the manuscript. This work was funded by the Estonian Research Council (PUT1317, 1399), the Swedish Research Council (Vetenskapsrådet, 2017–05019), and the European Regional Development Fund. SG acknowledges an NWO-VENI grant from the Netherlands Organization for Scientific Research (016. Veni.181.078).

AUTHORSHIP

All authors contributed to the design and performed the study, and FA performed data analysis and wrote the first draft of the manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ele.13904.

DATA AVAILABILITY STATEMENT

All metabarcoding data are available at NCBI's Sequence Read Archive (SRA) under accession number PRJEB19855. Other data and R codes necessary for the analysis are available at Zenodo repository (https://doi.org/10.5281/zenodo.5566446)

ORCID

Farzad Aslani https://orcid.org/0000-0003-0485-9800
Stefan Geisen https://orcid.org/0000-0003-0734-727X

REFERENCES

Allison, S.D. & Martiny, J.B. (2008) Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences*, 105, 11512–11519.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. & Huse, S.M. (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One*, 4, e6372.

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M. et al. (2018) Structure and function of the global topsoil microbiome. *Nature*, 560, 233–237.

Bahram, M., Kohout, P., Anslan, S., Harend, H., Abarenkov, K. & Tedersoo, L. (2016) Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *The ISME Journal*, 10, 885–896.

Bahram, M., Netherway, T., Hildebrand, F., Pritsch, K., Drenkhan, R., Loit, K. et al. (2020) Plant nutrient-acquisition strategies drive topsoil microbiome structure and function. *New Phytologist*, 227, 1189–1199.

Bastida, F., Eldridge, D.J., Abades, S., Alfaro, F.D., Gallardo, A., García-Velázquez, L. et al. (2020) Climatic vulnerabilities and ecological preferences of soil invertebrates across biomes. *Molecular Ecology*, 29, 752–761.

Bennett, J.R. & Gilbert, B. (2016) Contrasting beta diversity among regions: how do classical and multivariate approaches compare? *Global Ecology and Biogeography*, 25, 368–377.

Bie, T., Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D. et al. (2012) Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. *Ecology Letters*, 15, 740–747.

Briones, M.J.I. (2014) Soil fauna and soil functions: a jigsaw puzzle. *Frontiers in Environmental Science*, 2, 1–22. https://doi.org/10.3389/fenvs.2014.00007.

Caruso, T., Chan, Y., Lacap, D.C., Lau, M.C., McKay, C.P. & Pointing, S.B. (2011) Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. *The ISME Journal*, 5, 1406–1413.

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.

Chen, Q.-L., Hu, H.-W., Yan, Z.-Z., Li, C.-Y., Nguyen, B.-A.-T., Sun, A.-Q. et al. (2021) Deterministic selection dominates microbial community assembly in termite mounds. *Soil Biology and Biochemistry*, 152, 108073.

Crowther, T.W., Van den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A.D., Mo, L. et al. (2019) The global soil community and its influence on biogeochemistry. *Science*, 365, 772.

Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ, A. et al. (2015) Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science*, 349, 970–973.

Daws, S.C., Cline, L.A., Rotenberry, J., Sadowsky, M.J., Staley, C., Dalzell, B. et al. (2020) Do shared traits create the same fates? Examining the link between morphological type and the biogeography of fungal and bacterial communities. *Fungal Ecology*, 46, 100948.

de Groot, G.A., Laros, I. & Geisen, S. (2016) Molecular identification of soil eukaryotes and focused approaches targeting protist and faunal groups using high-throughput metabarcoding. In: Martin, F. & Uroz, S. (Eds.) *Microbial environmental genomics (MEG)*. New York, NY: Springer, Humana Press, pp. 125–140.

Decaëns, T. (2010) Macroecological patterns in soil communities. *Global Ecology and Biogeography*, 19, 287–302.

Delgado-Baquerizo, M., Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan, J.J., Encinar, D. et al. (2016) Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nature Communications*, 7, 10541.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D. et al. (2018) A global atlas of the dominant bacteria found in soil. *Science*, 359, 320–325.

Delgado-Baquerizo, M., Reich, P.B., Trivedi, C., Eldridge, D.J., Abades, S., Alfaro, F.D. et al. (2020) Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nature Ecology & Evolution*, 4, 210–220.

Dray, S., Legendre, P. & Blanchet, G. (2016) packfor: Forward selection with permutation (Canoco p. 46). R-Forge, The R project for statistical computing. Available at: http://R-Forge.R-project.org/projects/sedar/. Accessed 4 May 2016.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10, 996–998.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194–2200.

Evans, S., Martiny, J.B. & Allison, S.D. (2017) Effects of dispersal and selection on stochastic assembly in microbial communities. *The ISME Journal*, 11, 176–185.

Farjalla, V.F., Srivastava, D.S., Marino, N.A., Azevedo, F.D., Dib, V., Lopes, P.M. et al. (2012) Ecological determinism increases with organism size. *Ecology*, 93, 1752–1759.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L. et al. (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, 109, 21390–21395.

Fodelianakis, S., Valenzuela-Cuevas, A., Barozzi, A. & Daffonchio, D. (2021) Direct quantification of ecological drift at the population level in synthetic bacterial communities. *The ISME Journal*, 15, 55–66.

Forbes, A.E. & Chase, J.M. (2002) The role of habitat connectivity and landscape geometry in experimental zooplankton metacommunities. *Oikos*, 96, 433–440.

Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G. & Jousset, A. (2019) Protists: puppet masters of the rhizosphere microbiome. *Trends in Plant Science*, 24, 165–176.

Gaston, K.J. (2000) Global patterns in biodiversity. *Nature*, 405, 220–227.

Geisen, S., Mitchell, E.A.D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F. et al. (2018) Soil protists: a fertile frontier in soil biology research. *FEMS Microbiology Reviews*, 42, 293–323.

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M. et al. (2019) Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nature Communications*, 10, 1107.

Green, J.L., Bohannan, B.J. & Whitaker, R.J. (2008) Microbial biogeography: from taxonomy to traits. *Science*, 320, 1039–1043.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. & Martiny, J.B. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10, 497–506.

Hauswedell, H., Singer, J. & Reinert, K. (2014) Lambda: the local aligner for massive biological data. *Bioinformatics*, 30, i349–i355.

Hijmans, R. (2019) geosphere: Spherical Trigonometry. R package version 1.5–10. https://CRAN.R-project.org/package=geosphere.
Hildebrand, F., Tadeo, R., Voigt, A.Y., Bork, P. & Raes, J. (2014) LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome*, 2, 30.
Hillebrand, H. & Azovsky, A.I. (2001) Body size determines the strength of the latitudinal diversity gradient. *Ecography*, 24, 251–256.
Jia, X., Dini-Andreote, F. & Salles, J.F. (2018) Community assembly processes of the microbial rare biosphere. *Trends in Microbiology*, 26, 738–747.
Jombart, T. & Dray, S. (2008) Adephylo: exploratory analyses for the phylogenetic comparative method. *Bioinformatics*, 26, 1907–1909.
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.
Kinlock, N.L., Prowant, L., Herstoff, E.M., Foley, C.M., Akin-Fajiye, M., Bender, N. et al. (2018) Explaining global variation in the latitudinal diversity gradient: Meta-analysis confirms known patterns and uncovers new ones. *Global Ecology and Biogeography*, 27, 125–141.
Korhonen, J.J., Soininen, J. & Hillebrand, H. (2010) A quantitative analysis of temporal turnover in aquatic species assemblages across ecosystems. *Ecology*, 91, 508–517.
Ladau, J. & Eloe-Fadrosh, E.A. (2019) Spatial, temporal, and phylogenetic scales of microbial ecology. *Trends in Microbiology*, 27(8), 662–669. https://doi.org/10.1016/j.tim.2019.03.003
Lane, D.J. (1991) 16S/23S rRNA sequencing. In: Stackebrandt, E. & Goodfellow, M. (Eds.) *Nucleic acid techniques in bacterial systematics*. New York: John Wiley and Sons, pp. 115–175.
Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F. et al. (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7, 601–613.
Levins, R. (1968) *Evolution in changing environments*. Princeton, NJ: Princeton University Press.
Losos, J.B. (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, 11, 995–1003.
Luan, L., Jiang, Y., Cheng, M., Dini-Andreote, F., Sui, Y., Xu, Q. et al. (2020) Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. *Nature Communications*, 11, 6406.
Lynch, M.D. & Neufeld, J.D. (2015) Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology*, 13, 217–229.
Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L. et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4, 102–112.
Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. (2005) The microbial pan-genome. *Current Opinion in Genetics & Development*, 15, 589–594.
Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F. et al. (2013) Patterns and processes of microbial community assembly. *Microbiology and Molecular Biology Reviews*, 77, 342–356.
Nielsen, U.N., Osler, G.H., Campbell, C.D., Burslem, D.F. & van der Wal, R. (2010) The influence of vegetation type, soil properties and precipitation on the composition of soil mite and microbial communities at the landscape scale. *Journal of Biogeography*, 37, 1317–1328.
Ning, D., Deng, Y., Tiedje, J.M. & Zhou, J. (2019) A general framework for quantitatively assessing ecological stochasticity. *Proceedings of the National Academy of Sciences*, 116, 16892–16898.
Ning, D., Yuan, M., Wu, L., Zhang, Y., Guo, X., Zhou, X. et al. (2020) A quantitative framework reveals ecological drivers of grassland microbial community assembly in response to warming. *Nature Communications*, 11, 4717.
Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P. & McGlinn, D. (2020) vegan: community ecology package. R package version 2.5-7. https://CRAN.R-project.org/package=vegan
Oliverio, A.M., Geisen, S., Delgado-Baquerizo, M., Maestre, F.T., Turner, B.L. & Fierer, N. (2020) The global-scale distributions of soil protists and their contributions to belowground systems. Science. *Advances*, 6, eaax8787.
Phillips, H.R., Guerra, C.A., Bartz, M.L., Briones, M.J., Brown, G., Crowther, T.W. et al. (2019) Global distribution of earthworm diversity. *Science*, 366, 480–485.
Powell, J.R., Karunaratne, S., Campbell, C.D., Yao, H., Robinson, L. & Singh, B.K. (2015) Deterministic processes vary during community assembly for ecologically dissimilar taxa. *Nature Communications*, 6, 1–10.
Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J. et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35, 7188–7196.
Soininen, J., Korhonen, J.J. & Luoto, M. (2013) Stochastic species distributions are driven by organism size. *Ecology*, 94, 660–670.
Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313.
Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J. et al. (2013) Quantifying community assembly processes and identifying features that impose them. *The ISME Journal*, 7, 2069–2079.
Stegen, J.C., Lin, X., Konopka, A.E. & Fredrickson, J.K. (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal*, 6, 1653–1664.
Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I. et al. (2015) Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys*, 10, 1–43.
Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R. et al. (2014) Global diversity and geography of soil fungi. *Science*, 346(6213), 1078–1088. https://doi.org/10.1126/science.1256688.
van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D.A. et al. (2019) Soil nematode abundance and functional group composition at a global scale. *Nature*, 572, 194–198.
Vellend, M., Srivastava, D.S., Anderson, K.M., Brown, C.D., Jankowski, J.E., Kleynhans, E.J. et al. (2014) Assessing the relative importance of neutral stochasticity in ecological communities. *Oikos*, 123, 1420–1430.
Wagg, C., Bender, S.F., Widmer, F. & Van Der Heijden, M.G. (2014) Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proceedings of the National Academy of Sciences*, 111, 5266–5270.
Wickham, H. (2026) *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag.
Wickham, H., Romain, F., Lionel, H. & Kirill, M. (2020) dplyr: A Grammar of Data Manipulation. R package version 1.0.2. https://CRAN.R-project.org/package=dplyr.
Wilschut, R.A., van der Putten, W.H., Garbeva, P., Harkes, P., Konings, W., Kulkarni, P. et al. (2019) Root traits and belowground herbivores relate to plant–soil feedback variation among congeners. *Nature Communications*, 10, 1564.
Xing, D. & He, F. (2021) Analytical models for β-diversity and the power-law scaling of β-deviation. *Methods in Ecology and Evolution*, 12, 405–414.
Xiong, W.U., Jousset, A., Li, R., Delgado-Baquerizo, M., Bahram, M., Logares, R. et al. (2021) A global overview of the trophic

structure within microbiomes across ecosystems. *Environment International*, 151, 106438.
Xu, X., Wang, N., Lipson, D., Sinsabaugh, R., Schimel, J., He, L. et al. (2020) Microbial macroecology: in search of mechanisms governing microbial biogeographic patterns. *Global Ecology and Biogeography*, 29, 1870–1886.
Zhang, J.Y. (2016) Spaa: Species Association Analysis. R package version 0.2. 2. https://CRAN.R-project.org/package=spaa.
Zhou, J. & Ning, D. (2017) Stochastic community assembly: does it matter in microbial ecology? *Microbiology and Molecular Biology Reviews*, 81, e00002–17.
Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M. et al. (2019) Body size determines soil community assembly in a tropical forest. *Molecular Ecology*, 28, 528–543.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Aslani, F., Geisen, S., Ning, D., Tedersoo, L. & Bahram, M. (2022) Towards revealing the global diversity and community assembly of soil eukaryotes. *Ecology Letters*, 25, 65–76. https://doi.org/10.1111/ele.13904