Received: 12 January 2022 | Revised: 27 June 2022 | Accepted: 9 July 2022

DOI: 10.1111/1462-2920.16134

RESEARCH ARTICLE

environmental microbiology Applied Microbiology International

Hierarchical phylogenetic community assembly of soil protists in a temperate agricultural field

Julien Roy[1,2] | **Florent Mazel**[3] | **Kenneth Dumack**[4] | **Michael Bonkowski**[4] | **Matthias C. Rillig**[1,2]

[1]Institut für Biologie, Ökologie der Pflanzen, Freie Universität Berlin, Berlin, Germany

[2]Berlin–Brandenburg Institute of Advanced Biodiversity Research (BBIB), Berlin, Germany

[3]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

[4]Terrestrial Ecology Group, Institute of Zoology, University of Cologne, Cologne, Germany

Correspondence
Julien Roy, Freie Universität Berlin, Institute of Biology, Altensteinstr. 6, Berlin 14195, Germany.
Email: royjulien@zedat.fu-berlin.de

Funding information
Bundesministerium für Bildung und Forschung, Grant/Award Numbers: 031B0508B, 031B0515D; Federal Ministry of Education and Research

Abstract

Protists are abundant, diverse and perform essential functions in soils. Protistan community structure and its change across time or space are traditionally studied at the species level but the relative importance of the processes shaping these patterns depends on the taxon phylogenetic resolution. Using 18S rDNA amplicon data of the Cercozoa, a group of dominant soil protists, from an agricultural field in western Germany, we observed a turnover of relatively closely related taxa (from sequence variants to genus-level clades) across soil depth; while across soil habitats (rhizosphere, bulk soil, drilosphere), we observed turnover of relatively distantly related taxa, confirming Paracercomonadidae as a rhizosphere-associated clade. We extended our approach to show that closely related Cercozoa encounter divergent arbuscular mycorrhizal (AM) fungi across soil depth and that distantly related Cercozoa encounter closely related AM fungi across soil compartments. This study suggests that soil Cercozoa community assembly at the field scale is driven by niche-based processes shaped by evolutionary legacy of adaptation to conditions primarily related to the soil compartment, followed by the soil layer, giving a deeper understanding on the selection pressures that shaped their evolution.

INTRODUCTION

Protists are a functionally highly diverse paraphyletic group of ubiquitous unicellular eukaryotes (Geisen et al., 2018), which encompass most eukaryotic diversity on Earth (Burki et al., 2020). Their role in ecosystem functioning is increasingly recognized (Geisen et al., 2018), and as such protist distribution proved to be useful to tracking environmental changes, including past environments, and is considered to be an additional parameter relevant for biogeochemistry and climate models (Pawlowski et al., 2018). Community structure (occurrence and abundance of taxa within a community) and its change across time or space (e.g., betadiversity) are traditionally studied at species level. However, the relative importance of the processes shaping these patterns depends on the phylogenetic scale of taxon delimitation (Cavender-Bares et al., 2006; Chalmandrier et al., 2019; Groussin et al., 2017; Martiny et al., 2015; Mazel et al., 2017; Pillon et al., 2019; Roy et al., 2019). For example, if dispersal limitation has been important at different periods in the past, it would today manifest in extent taxa geographic distribution at different phylogenetic levels (e.g., disjunct distribution of higher clades if dispersal limitation was important in the deep past). In another example, if traits on which the current environment filters are conserved at different levels across the phylogeny (because of natural selection on specific traits at a given period), the correlation between environment and geographic distribution would vary across phylogenetic levels. Identifying those levels of phylotype resolution is

therefore of importance because it also subsequently influences the identification of the organisms associated with certain environments, or with certain states of the ecosystems (i.e., bioindicators). The phylogenetic scale dependency of betadiversity and its drivers can apply to multiple spatial and/or environmental gradients, and to any phylogenetic level from intraspecific diversity to deep evolutionary divergences. In summary, accounting for phylogenetic levels can reveal a hierarchical community assembly caused by multiple spatial and environmental drivers.

In protists, a number of traits with respect to morphology, locomotion or nutrition appear phylogenetically conserved (Dumack et al., 2020). Because traits can be conserved at different levels along the phylogenetic/taxonomic hierarchy, such as species, orders or genera (Lennon et al., 2012; Martiny et al., 2015), niche-based processes driven by different abiotic filters may all be more apparent at different levels of phylogenetic/taxonomic resolution (Groussin et al., 2017; Martiny et al., 2015; Roy et al., 2019). Protist community assembly probably follows this hierarchical pattern of community assembly: shift in protist functional composition, where functional traits are relatively phylogenetically conserved (Dumack et al., 2020), is observed across environmental gradients (e.g., Mazel et al., 2021), with entire clades restricted to particular environmental conditions (Fernández et al., 2022; Oliverio et al., 2020); while within genus, cryptic species can have little overlap in their realized niche (Singer et al., 2018) and even within species, different strains can specialize on different prey (Glücksman et al., 2010).

Phylogenetic scale is readily integrated in the study of bacterial (Martiny et al., 2009; Amend et al., 2016; Groussin et al., 2017; see Martiny et al., 2015 for a complete review), fungal (Chalmandrier et al., 2019; Roy et al., 2019), plant (Cavender-Bares et al., 2006) and animal (Mazel et al., 2017; Saladin et al., 2019) community ecology. The wealth of statistical methods that incorporate phylogenetic scale into community turnover analysis (Chalmandrier et al., 2015; Groussin et al., 2017; Martiny et al., 2013; Washburne et al., 2017) has rarely been applied to study protistan biogeography and community assembly processes. Recently, Lentendu and Dunthorn (2021) revealed a phylogenetic signal in operational taxonomic unit (OTU) co-occurrences and co-exclusions in terrestrial and marine protists. Their results suggest that environmental filtering and dispersal limitation are the dominant forces driving protist co-occurrences in both environments, whereas competitive exclusion is detected in the marine environment. Furthermore, the study of the community composition of soil protists along an elevation transect has proven that niche conservatism (probably in regards to thermal growth constraints) shapes their contemporary spatial distribution (Fernández et al., 2022). However, how phylogenetic relatedness and niche conservatism drives soil protist community assembly across multiple environmental gradients, whereby hierarchical phylogenetic community assembly is driven by functional traits that are conserved at different phylogenetic levels, is unknown.

Soil protists interact with a diverse community of soil microbes of various trophic levels and functional groups (Geisen et al., 2018). Besides predator–prey interactions between bacterivorous protists and bacteria, protists likely play important roles as fungi consumers (Dumack et al., 2018; Geisen et al., 2016). In addition, interactions between protists and arbuscular mycorrhizal (AM) fungi exist, such as the uptake by the fungus of nitrogen delivered from grazing of bacteria by protists (Henkes et al., 2018; Rozmos et al., 2022), or a flow of plant-assimilated carbon from AM fungi to protists (Hünninghaus et al., 2019). If these protists-fungi associations are somewhat specific, we expect cercozoan community composition to be correlated with fungal community composition. However, similarly to environmental gradients, whether there is a phylogenetic signal in the associations of protists with other microbes is unknown.

In this study, we first implemented a phylogenetic decomposition of betadiversity to measure the correlation between soil depth, soil compartment and the phylogenetic community structure of Cercozoa, a major clade of soil protists (Grossmann et al., 2016), in a single agricultural field in western Germany (Degrune et al., 2019) and we identified bioindicator cercozoan taxa at relevant phylogenetic resolutions. Second, we studied covariation in interkingdom betadiversity using AM fungal community data from the same samples (Sosa-Hernández et al., 2018) to reveal the phylogenetic breadth of co-occurring, and potentially interacting, taxa of the two microbial groups.

EXPERIMENTAL PROCEDURES

Dataset description

The cercozoan diversity in a single agricultural field in western Germany was characterized using Illumina MiSeq paired-end amplicon sequencing of the 18S rRNA V4 region (Degrune et al., 2019) using a nested-PCR with the forward primers S616F_Cerco and S616F_Eocer and the reverse primers S963R_Cerco and S947R Cerco (Fiore-Donno et al., 2018). The field is a plot experiment which has been performed at campus Klein–Altendorf near Bonn, Germany and was planted with *Cichorium intybus* L. at the time of sampling (Uksa et al., 2018). The soil is characterized as Haplic Luvisol. Soil samples were collected across two orthogonal environmental gradients. A horizontal gradient consisted of different soil compartments: the

drilosphere (defined as maximal 1 mm coating of earthworm holes), rhizosphere (maximal 2 mm root-adhering soil), and bulk soil. A vertical gradient of soil depth (hereafter referred to as layer), consisted of two levels: the topsoil layer (at 10–30 cm depth) and the subsoil layer (at 60–75 cm depth) in each of the soil compartments. Each combination (layer * compartment) was replicated three times, resulting in 18 communities.

Sequence variants inference

We used DADA2 in R (Callahan et al., 2016) to obtain denoised, chimaera-free, non-singleton cercozoan exact sequence variants (ESVs). Primers were removed. Forward and reverse sequences containing any ambiguous base were removed and trimmed to 220 and 270 bp respectively, to account for base quality. Sequences with more than 1 and 2 expected error, for the forward and reverse reads, respectively, were removed. Sequences were dereplicated and denoised into sequence variants using a model of nucleotide substitution error estimated from the data. Resulting paired-end sequences were merged. Chimaera were removed using the de novo approach in which ESVs that are a bimera of two other ESV subsets are considered chimaera. Cercozoan ESVs were identified against the PR2 database (Guillou et al., 2013) using the Naive Bayesian Classifier (Wang et al., 2007), at a minimum bootstrap confidence threshold of 100%.

Phylogeny reconstruction

We constructed a maximum-likelihood phylogenetic tree using reference cercozoan sequences and the inferred ESVs. Based on a BLAST search against the PR2, we obtained 164 full-length rRNA gene reference sequences closely related to the previously inferred cercozoan ESVs. These sequences were aligned using MAFFT (Katoh et al., 2002) to create a reference backbone alignment of long sequences. Cercozoan ESVs were aligned to this alignment using MAFFT to place short fragments onto aligned longer fragments. A maximum-likelihood phylogenetic tree was built in RAxML (Stamatakis, 2014) using all cercozoan ESVs and all reference sequences, by conducting a bootstrap analysis (100 bootstraps) under a GTRGAMMA model of nucleotide substitution.

Analysis of community-environment correlation across phylogenetic resolution

Prior to operational taxonomic unit (OTU) clustering and analysis of community-environment correlation across phylogenetic resolution, the ESV contingency table was rarefied to 32,000 reads to account for difference in sequencing depth among samples and obtain relative abundance. We used BDTT in R (Groussin et al., 2017) to cluster cercozoan ESVs into OTUs at genetic distance (in mean substitutions per site, subs/site) from the tips (0 subs/site) towards the root of the tree (up to 1.3 subs/site). BDTT maintains OTU monophyly in non-ultrametric trees. The abundance (or occurrence) of the new OTUs are the sum of read count (or incidences) of the ESVs that compose the new OTU. For each genetic distance, a new OTU contingency table was created and sample pairwise beta diversity was calculated using Bray–Curtis dissimilarity

We used non-parametric multivariate analysis of variance (Anderson 2001) using marginal test to calculate the unique part of variance in community dissimilarities explained by soil layer and soil compartment (i.e., the community-environment correlation). Statistical significance was assessed based on 999 Monte-Carlo permutations. To further test whether the community-environment correlation was the result of a phylogenetic signal and not solely attributable to the number of OTUs used to run the analysis, we randomized the tips of the phylogeny while keeping the community composition constant and we repeated the variance partitioning. A positive departure from the 95% null distribution of the community-environment correlation supports a significant phylogenetic signal. Community dissimilarities were further visualized using NMDS ordinations at representative phylogenetic resolutions. The statistical and phylogenetic analyses were conducted using the R packages vegan (Oksanen et al., 2016) and ape (Paradis et al., 2004).

To assess the robustness of our results and the potential for taxonomy in evaluating community-environment correlations across phylogenetic resolutions, we repeated the analyses using taxonomic binning of ESVs across taxonomic ranks. The analysis was conducted with all ESVs or restricting to ESVs with species level taxonomic annotation.

Identification of bioindicators

At the OTU phylogenetic resolution showing maximal correlation with the environmental gradients, we further identified the OTUs that were positively associated to each part of the gradient using the indicspecies R package (De Cáceres & Legendre, 2009).

Analysis of Cercozoa-AM fungi covariation

We also extend the phylogenetic decomposition of the beta diversity framework to the study of Cercozoa-AM fungi covariation. We used AM fungal community data from the same samples as used for Cercozoa (Sosa-

Hernández et al., 2018). The ESV and phylogenetic tree inference for the AM fungal data followed a similar analytical strategy as described above for Cercozoa (Roy et al., 2019). We analysed the covariation of beta diversity (Bray–Curtis dissimilarity) patterns of Cercozoa and AM fungi using Mantel tests (Legendre & Legendre, 2012). The Mantel correlation (Spearman coefficient) was assessed at each OTU genetic distance for both AM fungi and Cercozoa.

RESULTS

We inferred 1454 cercozoan ESVs totalling 576,000 reads after random read subsampling to 32,000 reads per sample. Clustering ESVs into broader OTUs revealed that most taxonomic diversity was within OTUs delineated at 0.1 subs/site with only small subsequent changes in OTU numbers as genetic distance increased (Figure 1A).

Community-environment correlation across phylogenetic resolution

At the ESV level (0 subs/site, i.e., the taxonomically most highly resolved tree) soil layer (topsoil vs. subsoil), and soil compartment (rhizosphere, drilosphere, bulk soil) explained 20% and 16% of community composition (Figure 1B). However, the correlation of community composition with soil layer and soil compartment strongly varied with OTU phylogenetic resolution. This result was consistent using taxonomic ranks instead of phylogenetic distance, restricting or not to ESVs with species annotation (Figure S1). The community correlation with the soil layer remained constant up to a resolution of 0.5 subs/site (or genus level using taxonomic binning; Figure S1), but fell within the null envelope generated by phylogenetic randomization, indicating that this result could not be attributed to a phylogenetic signal (Figure 1B). These results indicate that beta diversity patterns were robust to change in phylogenetic resolution up to 0.5 subs/site, but probably based on highly influential ESVs driving the pattern. After 0.5 subs/site (or the genus level), the community correlation with soil layer dropped. In contrast, the community correlation with the soil compartment increased constantly up to 1.1 subs/site (family to order level using taxonomic binning; Figure S1). The community-environment correlation at 1.1 subs/site genetic distance marked a significant threshold that could be attributed to a phylogenetic signal (Figure 1B). Our results are well exemplified in non-metric multidimensional scaling (NMDS) ordinations: it show the clustering of communities by soil depth to clustering by soil compartment as phylogenetic resolution decreases, especially a convergence of topsoil and subsoil communities in the rhizosphere (Figure 1C).

Bioindicator clades across the phylogenetic resolution

We found two abundant clades at 1.1 subs/site as being indicators of either the drilosphere or the rhizosphere soil compartments (Figure 2). The rhizosphere-associated clade comprised 150 ESVs mostly distributed within the genus *Paracercomonas* (Figure 2B). This rhizosphere-associated clade had a relative abundance of ~12% in the drilosphere and in bulk soil but reached ~20% in the rhizosphere. In contrast, the clade associated with the drilosphere comprised more than 1000 ESVs, belonging to multiple orders. It reached a relative abundance of ~75% in the drilosphere but dropped to ~62% in the rhizosphere. We also observed a very low-abundance clade, composed of three ESVs closely related to the rhizosphere-associated clade and annotated to *Paracercomonas* sp. (Figure 2A, B), absent in the rhizosphere, and reaching ~0.05% in the bulk soil, peaking sometimes at 0.25% (e.g., in the subsoil) (Figure 2C).

Co-variation between AM fungi and Cercozoa beta diversity across phylogenetic resolution

We further characterized the phylogenetic resolution of covariation in community turnover between Cercozoa and AM fungi. We identified the strongest correlation at the ESV level for Cercozoa, and at a resolution of 0.02 subs/site for AM fungi (Figure 3). In general, we identified a disparity of phylogenetic scale in their covariation: the composition of relatively closely related cercozoan taxa tended to covary with the composition of relatively distantly related AM fungi.

DISCUSSION

Community structure and its change across time or space are traditionally studied at species level. However, the relative importance of the processes shaping these patterns depends on the phylogenetic level of taxon delimitation, because dispersal limitation or natural selection on specific traits could have been important or happening at different periods in the past (and thus manifest in extent taxa geographic distribution at different phylogenetic levels, and/or in traits being conserved at different levels across the phylogeny). Identifying those phylotype resolutions is therefore of importance to

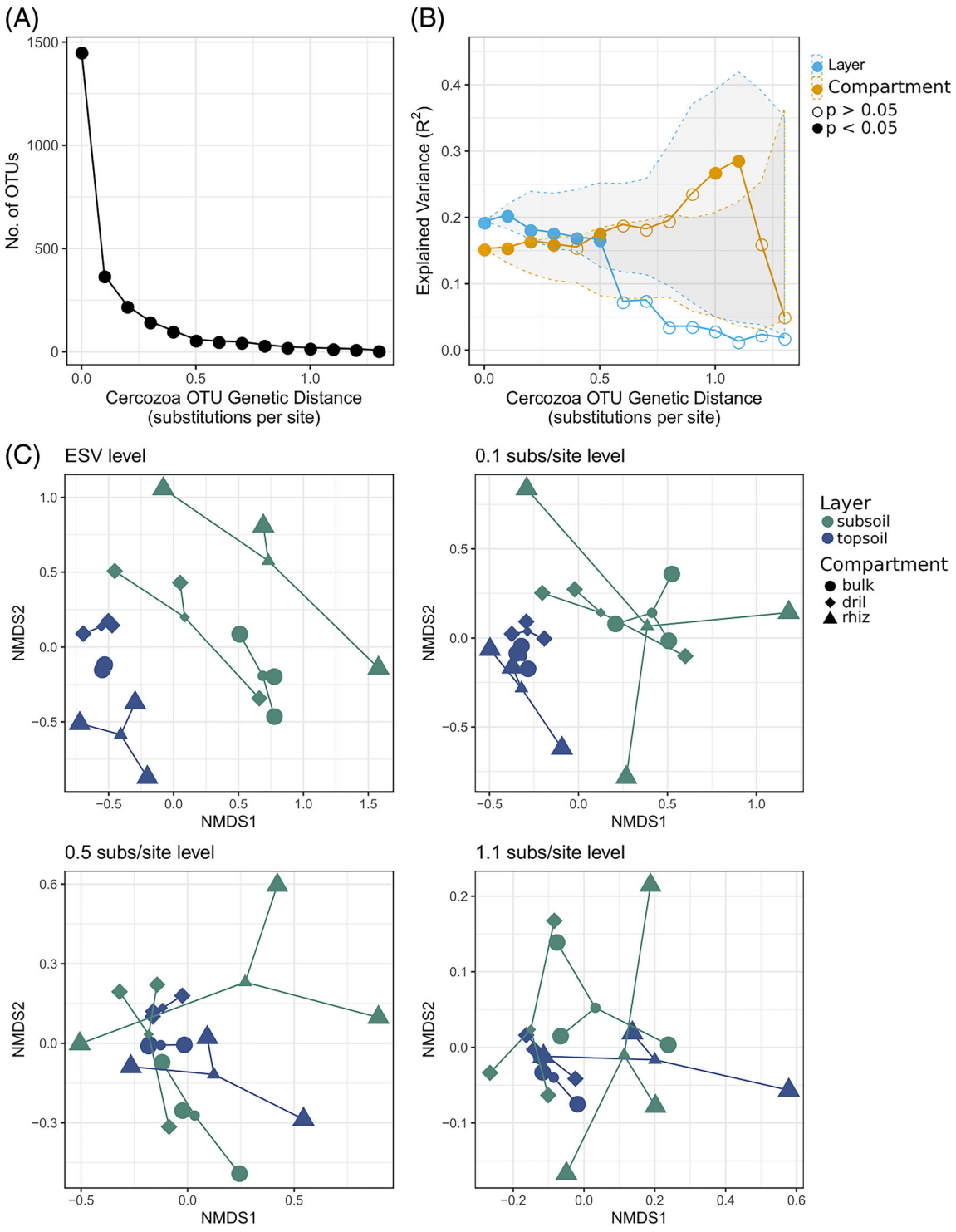

FIGURE 1 Community-environment correlation across phylogenetic depth and environmental gradients. (A) Number of operational taxonomic unit (OTUs) as a function of OTU genetic distance. (B) Community-environment correlation (variance explained, R2) with soil layer (blue) and soil compartment (orange) as a function of OTU genetic distance. The grey shaded area represents the 5%–95% percentiles of the community-environment correlation when randomizing phylogenetic relationships. (C) Non-metric multidimensional scaling (NMDS) ordinations of Cercozoa communities based on Bray–Curtis dissimilarity at four OTU genetic distances. Subs/site: Genetic distance in mean substitution per site. Sub: Subsoil; top: Topsoil; bulk: Bulk soil; dril: Drilosphere; rhiz: Rhizosphere

unravelling which and how abiotic environmental factors influence the distribution and community composition of protists, what is the most meaningful taxonomic levels to study the diversity and functioning of soil protists, what is the link between phylogeny and ecological function, and what are the organisms associated with certain

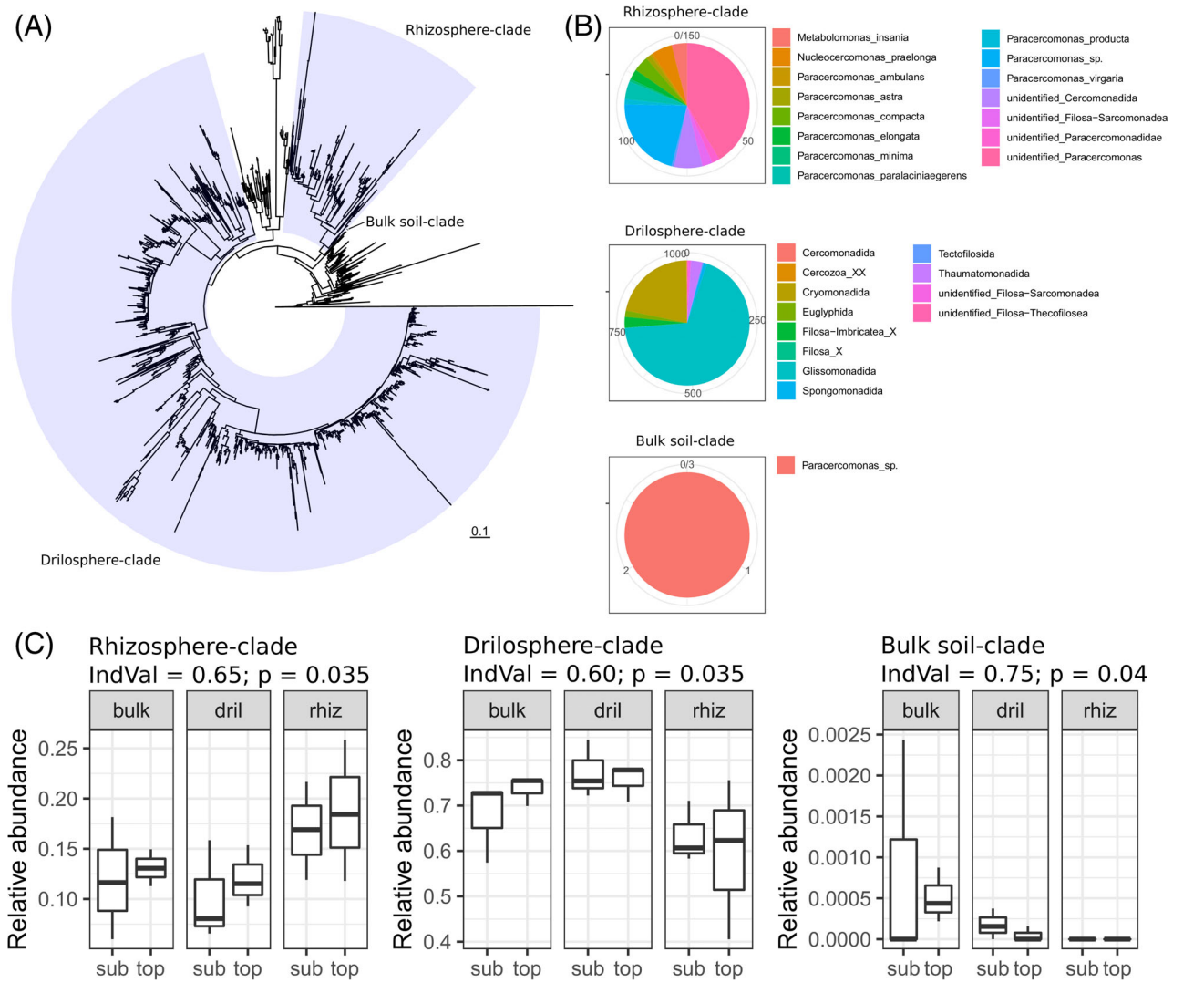

FIGURE 2 Bioindicator clades of Cercozoa in relation to soil compartment. (A) Maximum-likelihood phylogenetic tree of cercozoan ESVs depicting relationships between clades associated to different soil compartments. Three clades at 1.1 substitution per site and with a significant association to either rhizosphere, drilosphere or bulk soil are colour shaded. Scale bar: 0.1 mean substitution per site. (B) Pie charts represent the class and species-level composition (ESV richness) for the clades associated with the rhizosphere, drilosphere or bulk soil, respectively. (C) Relative abundance of the three clades at 1.1 substitution per site with a significant association to either rhizosphere, drilosphere or bulk soil. Sub: Subsoil; top: Topsoil; bulk: Bulk soil; dril: Drilosphere; rhiz: Rhizosphere

environments, or with certain states of the ecosystems (i.e., bioindicators) (Caron & Hu, 2018; Geisen et al., 2017).

Phylogenetic structure across soil depth and habitat reveals a hierarchical community assembly of Cercozoa at the field scale

We observed a turnover of relatively closely related taxa (from sequence variants to genus-level clades) across soil depths but a turnover of relatively distantly related taxa (above the genus level up to order level) across soil habitats. This presents similarity with Humboldt's observations on worldwide floras (Humboldt and Bonpland 1814) as recently confirmed for ultramafic island's flora (Pillon et al., 2019) but also for AM fungi in agricultural fields (Roy et al., 2019; Roy et al., 2021), mammal gut bacteria (Groussin et al., 2017; Martiny et al., 2015) and birds and mammals at global scales (Mazel et al., 2017). This suggests that soil Cercozoa community assembly at the field scale is driven by niche-based processes shaped by evolutionary legacy of adaptation to conditions primarily related to soil compartment, followed by soil layer.

Abiotic conditions in different soil compartments and at different soil depths, respectively, could have driven the early and more recent evolution of Cercozoa. In

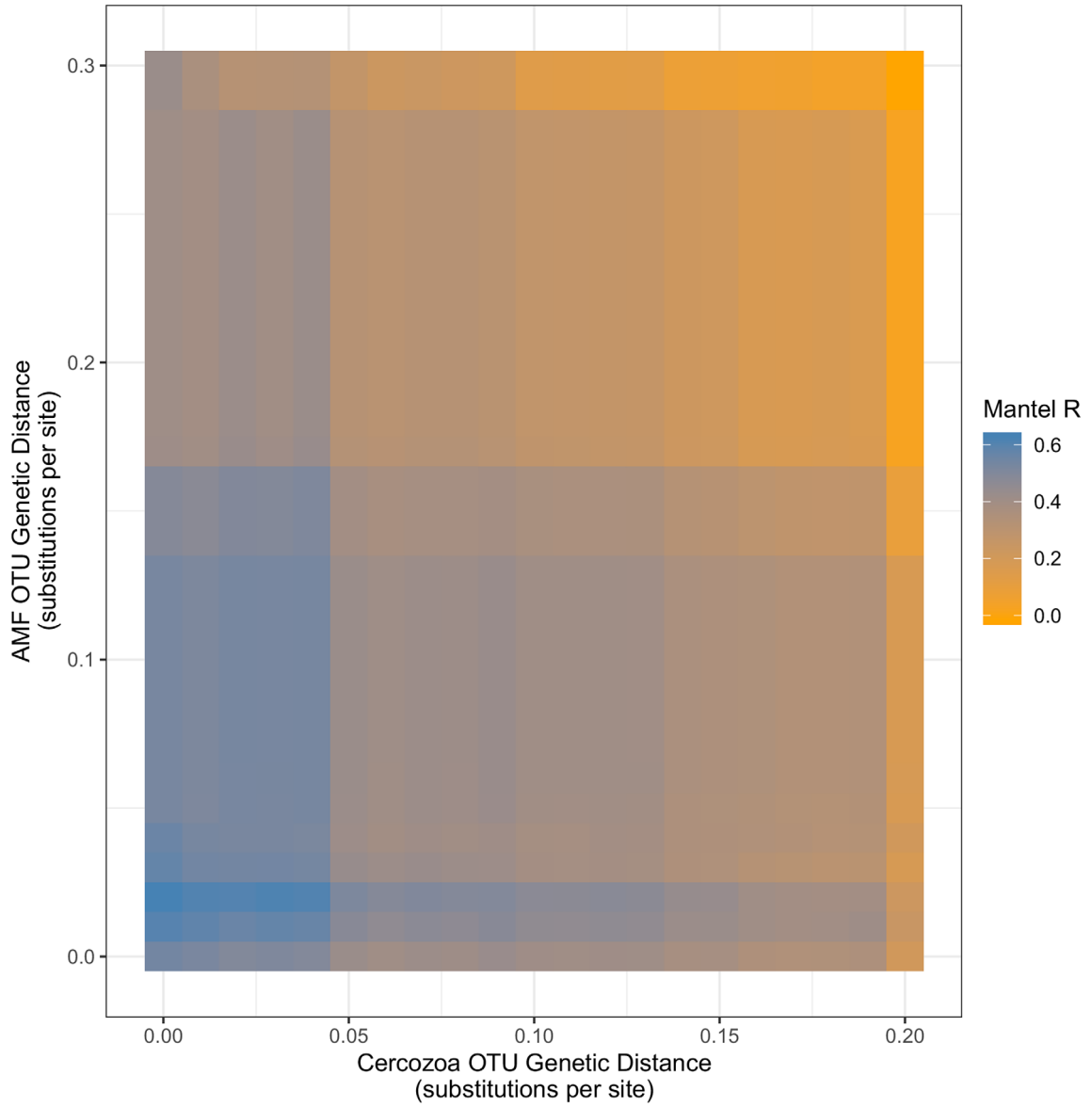

FIGURE 3 Covariation between cercozoan and AM fungal beta diversity. The covariation between was assessed using Mantel Pearson's R correlation across operational taxonomic unit (OTU) genetic thresholds for both Cercozoa and AM fungi

contrast, the soil compartment was shown to have little influence on AM fungi (Roy et al., 2019; Sosa-Hernández et al., 2018). The abiotic (e.g., soil water content or compaction, carbon and nitrogen concentration) and biotic (e.g., other microorganism abundance) factors of different soil compartments might have been of strong selective pressures in the evolution of Cercozoa, whereas they could have been of less importance for AM fungi probably due to the mycelium nature of AM fungi propagating from the rhizosphere to surrounding bulk soil, or efficient spore dispersal. Conversely to Cercozoa, we observed AM fungi of deeply divergent clades across soil depths (Roy et al., 2019). Differences in abiotic conditions across soil depths could impose strong constraints on AM fungi (Sosa-Hernández et al., 2018).

Potential for the ESV approach in Cercozoa metabarcoding studies

Similarly to other microbes such as AM fungi (Hart & Reader, 2002; Lekberg et al., 2014; Powell & Sikes, 2014; Roy et al., 2019) and bacteria (Philippot et al., 2010; Martiny et al., 2015), our results support the notion of ecological coherence in deep phylogenetic branches in Cercozoa. There was no additional benefit of increased phylogenetic resolution below the genus level to study the change in cercozoan community structure along the environmental gradients studied here (soil depth and compartment). Sequence variant resolution revealed similar patterns as observed at higher phylogenetic level (this study) or OTUs binned with other methods (Degrune et al., 2019). This is

consistent with previous studies showing no or weak influence of taxonomic resolution to infer protist palaeoecological dynamics (Mitchell et al., 2014), or other studies on soil fungi showing relatively robust betadiversity patterns to OTU delimitation methods and genetic thresholds (Botnen et al., 2018; Glassman et al., 2018; Roy et al., 2019). However, this is not always the case, as the ESV level can be too variable (Roy et al., 2021) or reveal additional drivers (Roy et al., 2019). Rather, our results indicate that broadening phylogenetic resolution may strengthen niche-related signals (Lu et al., 2016; Roy et al., 2019, 2021), at least up to a certain phylogenetic level. This can be explained by removing uncertainty in quantifying spatio-temporal distributions of finely resolved taxa. Such uncertainty arises from either stochastic processes, or deterministic processes imposed by another (unmeasured) axis of environmental variation. Indeed, closely related species can have little overlap in their realized niche, due to resource partitioning and a strong influence of environmental filtering when considering additional environmental gradients and/or spatial scales, as evidenced among cryptic species of testate (shelled) amoeba within the genus *Nebela* in relation to soil N content and water table depth (Singer et al., 2018). ESV resolution can therefore reveal niche partitioning or dispersal limitation as spatial extent increases and spatial grain decreases, such as centimetre-scale gradient within soil or sampling contrasting habitats and/or biogeographic provinces. From methodological considerations, because ESVs have advantages of increasing comparability among studies (Callahan et al., 2017) and the potential to resolve intraspecific diversity (Callahan et al., 2016), we advocate these sequences to be deposited even if further clustering is performed for community analyses. However, whether ESV approaches have the potential to resolve cryptic protistan species needs further research, and partial 18S as used in most protistan metabarcoding studies might still not have sufficient resolution to discriminate between cryptic species.

Specialization of Paracercomonadidae to the rhizosphere

Indicator analysis showed that the differences in community composition between soil compartments were mostly due to a single clade, Paracercomonadidae. Paracercomonadidae were more abundant in the rhizosphere than in any other soil compartment, irrespective of soil depth. The Paracercomonadidae were composed of three different genera, *Paracercomonas*, *Nucleocercomonas* and *Metabolomonas*, of which *Paracercomonas* was by far the most diverse. Members of *Paracercomonas* in fact have been often reported to be abundant in the rhizosphere of various plants (Flues et al., 2018; Sapp et al., 2018; Simonin et al., 2020) suggesting this is a highly specialized group in the plant rhizosphere. Paracercomonadidae are naked amoeboflagellate, gliding, bacterivores but the traits underlying this specialization and the specific functions performed by these organisms in the rhizosphere remain enigmatic. Selection may act on metabolic traits, fine-scale food preferences, susceptibility to parasites, the presence of predators in certain soil compartments or the production of different toxins (Caron & Hu, 2018; Öztoprak et al., 2020; Singer et al., 2018).

The phylogenetic resolution of interkingdom co-variation

Covariation in space between Cercozoa and AM fungi indicates that clades from each group tend to co-occur non-randomly. Interactions between AM fungi and protists are relatively diffuse (Henkes et al., 2018; Hünninghaus et al., 2019; Rozmos et al., 2022). It seems unlikely that such interactions may become specialized between co-occurring Cercozoa and AM fungi. Rather, non-random co-occurrences are likely the results of shared habitat between clades of the two groups. However, the analysis of Cercozoa-AM fungi covariation across phylogenetic resolution can generate hypotheses regarding interactions between soil protists and other soil organisms (Chalmandrier et al., 2019; Geisen et al., 2017), as the organisms that co-occur will be in contact, and hypothetically interact, either positively or negatively. In particular, the covariation of cercozoan species with larger AM fungal clades suggests that closely related Cercozoa will encounter strongly divergent AM fungi across soil depth. Inversely, across soil compartments, distantly related Cercozoa will encounter closely related AM fungi. From the fungal perspective, individuals (ramets) of the same fungal population are potentially in contact (or interacting) with distantly related, putatively functionally divergent, Cercozoa in the different soil compartments, or even the same AM fungus (within a defined soil layer) in the different compartments where its mycelium forages. In our case study, members of the AM fungal family Glomeraceae were shown to preferentially associate with and dominate in the rhizosphere at this field site (Sosa-Hernández et al., 2018), a clade that preferentially colonizes roots and does not extend far in soil (Hart & Reader, 2002; Powell et al., 2009), probably not reaching other compartments than the rhizosphere. Thus, the rhizosphere is a particular habitat where members of an entire clade, the Paracercomonadidae, co-occur, and potentially interact, with members of an entire family of AM fungi, the Glomeraceae. However, we note that phylogenetic distances are not directly comparable between the two taxa (i.e., Glomeromycota AM fungi or Cercozoa) and the

phylogenetic level disparity may actually reflect comparable evolutionary levels.

Phylogenetic relatedness and taxonomic binning are of equivalent utility for unravelling the community assembly processes of Cercozoa

We observed equivalent potential of phylogenetic relatedness and taxonomic binning to study community-environment correlation across the evolutionary history of Cercozoa. Our study complements recent results for protists (Lentendu & Dunthorn, 2021), and former results for fungi (Roy et al., 2019), showing equivalence between phylogenetically-based distances and pairwise sequence-based distances. This is not surprising. Since the early 2000s, cercozoan systematics have been deeply revised based on molecular phylogenies, so that paraphyly due to conflict between earlier morphological taxonomy and more recent advent of phylogenetic reconstruction of evolutionary relationships should not be common. Genera are defined on a molecular basis, and possible synapomorphies have been searched *a posteriori*. On the one hand, sequence binning using phylogenetic relatedness could still be more suitable than taxonomic binning for poorly known organisms such as protists, where missing taxonomic information below the phylum level is common, and new species are expected to be discovered (Bass et al., 2009; Howe et al., 2011). On the other hand, environmental DNA diversity surveys are usually based on short DNA markers, which are not fully suitable for robust phylogenetic inference.

Conclusions

We observed a spatial turnover of relatively closely related protists (from partial 18S sequence variants to genus-level clades) across the soil depth profile but a spatial turnover of relatively distantly related taxa (above the genus level up to order level) across soil habitats (bulk soil, rhizosphere, drilosphere). For example, we confirmed that an entire large clade, Paracercomonadidae is rhizosphere-associated. This suggests that soil Cercozoa community assembly at the field-scale is driven by niche-based processes shaped by evolutionary legacy of adaptations to conditions in relation primarily to soil compartment and followed by soil layer. We showed no additional benefit of increased phylogenetic resolution below the genus level (and therefore of the ESV level compared to OTUs at 97% sequence similarity) to study cercozoan betadiversity along the environmental gradients studied but this could change when considering other environmental gradients or spatial scales. Furthermore, closely-related Cercozoa segregated across soil depth may encounter strongly divergent AM fungi, whereas across soil compartments, members of the same fungal population, or the same fungus, may encounter strongly divergent Cercozoa. However, the rhizosphere in this field was a particular habitat where members of an entire cercozoan lineage encountered members of an entire AM fungal lineage.

AUTHOR CONTRIBUTIONS

Julien Roy developed the original idea of the study, performed data analyses and wrote the manuscript. Matthias C. Rillig and Michael Bonkowski contributed to funding acquisition. Florent Mazel, Kenneth Dumack, Michael Bonkowski and Matthias C. Rillig contribute to ideas and manuscript editing. All authors contributed to revision and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

This research was financed by the German Federal Ministry of Education and Research (BMBF) initiative 'BonaRes—Soil as a sustainable resource for the bioeconomy' within the 'INPLAMINT—Increasing agricultural nutrient-use efficiency by optimizing plant-soil-microorganisms interactions' project (grant number: 031B0508B), as well as by the project Soil3 in the same program (grant number: 031B0515D). We acknowledge comments of two anonymous reviewers which improved previous versions of the manuscript. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

DATA AVAILABILITY STATEMENT

Raw fastq sequence files of Cercozoa are accessible at the European Nucleotide Archive (ENA) under study accession number PRJEB30791. The analysed cercozoan dataset including the ESV count table, ESV taxonomy, and sample metadata is available as a single .RDS R object at figshare (10.6084/m9.figshare.20331681). Raw fastq sequence files of Glomeromycota are accessible at the European Nucleotide Archive (ENA) under study accession number PRJEB54814. The analysed glomeromycotan dataset including the ESV count table, ESV taxonomy, and sample metadata is available as a single. RDS R object at figshare (10.6084/m9.figshare.20337024).

ORCID

Julien Roy https://orcid.org/0000-0003-2964-1314
Florent Mazel https://orcid.org/0000-0003-0572-9901
Kenneth Dumack https://orcid.org/0000-0001-8798-0483
Michael Bonkowski https://orcid.org/0000-0003-2656-1183

Matthias C. Rillig https://orcid.org/0000-0003-3541-7853

REFERENCES

Amend, A.S., Martiny, A.C., Allison, S.D., Berlemont, R., Goulden, M. L., Lu, Y. et al. (2016) Microbial response to simulated global change is phylogenetically conserved and linked with functional potential. *The ISME Journal*, 10, 109–118.

Anderson, M.J. (2001) A new method for non parametric multivariate analysis of variance. *Austral Ecology*, 26, 32–46.

Bass, D., Howe, A.T., Mylnikov, A.P., Vickerman, K., Chao, E.E., Edwards Smallbone, J. et al. (2009) Phylogeny and classification of Cercomonadida (Protozoa, Cercozoa): Cercomonas, Eocercomonas, Paracercomonas, and Cavernomonas gen. nov. *Protist*, 160, 483–521.

Botnen, S.S., Davey, M.L., Halvorsen, R. & Kauserud, H. (2018) Sequence clustering threshold has little effect on the recovery of microbial community structure. *Molecular Ecology Resources*, 1, 13.

Burki, F., Roger, A.J., Brown, M.W. & Simpson, A.G.B. (2020) The new tree of eukaryotes. *Trends in Ecology & Evolution*, 35, 43–55.

Callahan, B.J., McMurdie, P.J. & Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, 11, 2639–2643.

Callahan, B.J., Mcmurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A. J. & Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581–583.

Caron, D.A. & Hu, S.K. (2018) Are we overestimating protistan diversity in nature ? *Trends in Microbiology*, 27, 197–205.

Cavender-Bares, J., Keen, A. & Miles, B. (2006) Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology*, 87, S109–S122.

Chalmandrier, L., Münkemüller, T., Lavergne, S. & Thuiller, W. (2015) Effects of species' similarity and dominance on the functional and phylogenetic structure of a plant meta-community. *Ecology*, 96, 143–153.

Chalmandrier, L., Pansu, J., Zinger, L., Boyer, F., Coissac, E., Génin, A. et al. (2019) Environmental and biotic drivers of soil microbial β-diversity across spatial and phylogenetic scales. *Ecography*, 42, 1–13.

De Cáceres, M. & Legendre, P. (2009) Associations between species and groups of sites: indices and statistical inference. *Ecology*, 90, 3566–3574.

Degrune, F., Dumack, K., Fiore-Donno, A.M., Bonkowski, M., Sosa-Hernández, M.A., Schloter, M. et al. (2019) Distinct communities of Cercozoa at different soil depths in a temperate agricultural field. *FEMS Microbiol Ecol*, 95, 1–7.

Dumack, K., Fiore-Donno, A.M., Bass, D. & Bonkowski, M. (2020) Making sense of environmental sequencing data: ecologically important functional traits of the protistan groups Cercozoa and Endomyxa (Rhizaria). *Molecular Ecology Resources*, 20, 398–403.

Dumack, K., Pundta, J. & Bonkowski, M. (2018) Food choice experiments indicate selective Fungivorous predation in *Fisculla terrestris* (Thecofilosea, Cercozoa). *J Eukaryot Microbiol E*, 66, 525–527.

Fernández, L.D., Seppey, C.V.W., Singer, D., Fournier, B., Tatti, D., Mitchell, E.A.D. et al. (2022) Niche conservatism drives the Elevational diversity gradient in major groups of free-living soil unicellular eukaryotes. *Microbial Ecology*, 83, 459–469.

Fiore-Donno, A.M., Rixen, C., Rippin, M., Glaser, K., Samolov, E., Karsten, U. et al. (2018) New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput environmental diversity surveys, with emphasis on worldwide biological soil crusts. *Molecular Ecology Resources*, 18, 229–239.

Flues, S., Blokker, M., Dumack, K. & Bonkowski, M. (2018) Diversity of *Cercomonad* species in the phyllosphere and rhizosphere of different plant species with a description of *Neocercomonas epiphylla* (Cercozoa, Rhizaria) a leaf-associated protist. *The Journal of Eukaryotic Microbiology*, 65, 587–599.

Geisen, S., Koller, R., Hünninghaus, M., Dumack, K., Urich, T. & Bonkowski, M. (2016) The soil food web revisited: diverse and widespread mycophagous soil protists. *Soil Biology and Biochemistry*, 94, 10–18.

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.

Geisen, S., Mitchell, E.A.D., Wilkinson, D.M., Adl, S., Bonkowski, M., Brown, M.W. et al. (2017) Soil protistology rebooted: 30 fundamental questions to start with. *Soil Biology and Biochemistry*, 111, 94–103.

Glassman, S.I. & Martiny, B.H. (2018) Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *mSphere*, 3, 1–5.

Glücksman, E., Bell, T., Griffiths, R.I. & Bass, D. (2010) Closely related protist strains have different grazing impacts on natural bacterial communities. *Environmental Microbiology*, 12, 3105–3113.

Grossmann, L., Jensen, M., Heider, D., Jost, S., Glücksman, E., Hartikainen, H. et al. (2016) Protistan community analysis: key findings of a large-scale molecular sampling. *The ISME Journal*, 10, 2269–2279.

Groussin, M., Mazel, F., Sanders, J.G., Smillie, C.S., Lavergne, S., Thuiller, W. et al. (2017) Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nature Communications*, 8, 1–12.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L. et al. (2013) The Protist ribosomal reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41, 597–604.

Hart, M.M. & Reader, R.J. (2002) Taxonomic basis for variation in the colonization strategy of arbuscular mycorrhizal fungi. *The New Phytologist*, 153, 335–344.

Henkes, G.J., Kandeler, E., Marhan, S., Scheu, S. & Bonkowski, M. (2018) Interactions of mycorrhiza and protists in the rhizosphere systemically alter microbial community composition, plant shoot-to-root ratio and within-root system nitrogen allocation. *Front Environ Sci*, 6, 117. https://doi.org/10.3389/fenvs.2018.00117

Howe, A.T., Bass, D., Scoble, J.M., Lewis, R., Vickerman, K., Arndt, H. et al. (2011) Novel cultured protists identify deep-branching environmental DNA clades of Cercozoa: new genera Tremula, Micrometopion, Minimassisteria, Nudifila, Peregrinia. *Annals of Anatomy*, 162, 332–372.

Hünninghaus, M., Dibbern, D., Kramer, S., Koller, R., Pausch, J., Schloter-Hai, B. et al. (2019) Disentangling carbon flow across microbial kingdoms in the rhizosphere of maize. *Soil Biology and Biochemistry*, 134, 122–130.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30, 3059–3066.

Legendre, P. & Legendre, L. (2012) *Numerical ecology*. Amsterdam: Elsevier Science BV.

Lekberg, Y., Gibbons, S.M. & Rosendahl, S. (2014) Will different OTU delineation methods change interpretation of arbuscular mycorrhizal fungal community patterns? *New Phytology*, 202, 1101–1104.

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.

Lentendu, G. & Dunthorn, M. (2021) Phylogenetic relatedness drives protist assembly in marine and terrestrial environments. *Global Ecology and Biogeography*, 30, 1532–1544.

Lu, H., Yeh, Y., Sastri, A.R., Shiah, F., Gong, G. & Hsieh, C. (2016) Evaluating community—environment relationships along fine to broad taxonomic resolutions reveals evolutionary forces underlying community assembly. *The ISME Journal*, 10, 2867–2878.

Martiny, A.C., Tai, A.P.K., Veneziano, D., Primeau, F. & Chisholm, S. W. (2009) Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. *Environmental Microbiology*, 11, 823–832.

Martiny, A.C., Treseder, K. & Pusch, G. (2013) Phylogenetic conservatism of functional traits in microorganisms. *The ISME Journal*, 7, 830–838.

Martiny, J.B.H., Jones, S.E., Lennon, J.T. & Martiny, A.C. (2015) Microbiomes in light of traits: a phylogenetic perspective. *Science*, 350, aac9323-1–aac9323-8.

Mazel, F., Malard, L., Niculita-Hirzel, H., Yashiro, E., Mod, H.K., Mitchell, E.A.D. et al. (2021) Soil protist function varies with elevation in the Swiss Alps. *Environmental Microbiology*, 24, 1689–1702. https://doi.org/10.1111/1462-2920.15686

Mazel, F., Wüest, R.O., Lessard, J.-P., Renaud, J., Ficetola, G.F., Lavergne, S. et al. (2017) Global patterns of β-diversity along the phylogenetic time-scale: the role of climate and plate tectonics. *Global Ecology and Biogeography*, 26, 1211–1221.

Mitchell, E.A.D., Lamentowicz, M., Payne, R.J. & Mazei, Y. (2014) Effect of taxonomic resolution on ecological and palaeoecological inference e a test using testate amoeba water table depth transfer functions. *Quaternary Science Reviews*, 91, 62–69.

Oksanen, J.F., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. et al. (2016) vegan: Community Ecology Package.

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, 1–11.

Öztoprak, H., Walden, S., Heger, T., Bonkowski, M. & Dumack, K. (2020) What drives the diversity of the most abundant terrestrial Cercozoan family (Rhogostomidae, Cercozoa, Rhizaria)? *Microorganisms*, 8, 1–17.

Paradis, E., Claude, J. & Strimmer, K. (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A. et al. (2018) The future of biotic indices in the ecogenomic era: integrating (e) DNA metabarcoding in biological assessment of aquatic ecosystems. *Sci Total Environ*, 637–638, 1295–1310.

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.

Pillon, Y., Alfonso, D., Herizo, G., Ii, P.P.L., Jaffré, T. & Merlot, S. (2019) Parallel ecological filtering of ultramafic soils in three distant Island floras. *Journal of Biogeography*, 46, 2457–2465.

Powell, J.R., Parrent, J.L., Hart, M.M., Klironomos, J.N., Rillig, M.C. & Maherali, H. (2009) Phylogenetic trait conservatism and the evolution of functional trade-offs in arbuscular mycorrhizal fungi. *Proc R Soc Biol Sci*, 276, 4237–4245.

Powell, J.R. & Sikes, B.A. (2014) Method or madness: does OTU delineation bias our perceptions of fungal ecology? *The New Phytologist*, 202, 1095–1097.

Roy, J., Mazel, F., Sosa-Hernandez, A., Duenas, J.F., Hempel, S., Zinger, L. et al. (2019) The relative importance of ecological drivers of arbuscular mycorrhizal fungal distribution varies with taxon phylogenetic resolution. *The New Phytologist*, 224, 936–948.

Roy, J., van Duijnen, R., Leifheit, E.F., Mbedi, S., Temperton, V.M. & Rillig, M.C. (2021) Legacy effects of pre-crop plant functional group on fungal root symbionts of barley. *Ecological Applications*, 31, 1–16.

Rozmos, M., Bukovská, P., Hršelová, H., Kotianová, M., Dudáš, M., Gančarčíková, K. & Jansa, J. (2022) Organic nitrogen utilisation by an arbuscular mycorrhizal fungus is mediated by speci fi c soil bacteria and a protist. *ISME Journal*, 16, 676–685.

Saladin, B., Thuiller, W., Graham, C.H., Lavergne, S., Maiorano, L., Salamin, N. et al. (2019) Environment and evolutionary history shape phylogenetic turnover in European tetrapods. *Nature Communications*, 10, 1–9.

Sapp, M., Ploch, S., Fiore-donno, A.M., Bonkowski, M. & Rose, L.E. (2018) Protists are an integral part of the Arabidopsis thaliana microbiome. *Environmental Microbiology*, 20, 30–43.

Simonin, M., Dasilva, C., Terzi, V., Ngonkeu, E.L.M., Diouf, D., Kane, A. et al. (2020) Influence of plant genotype and soil on the wheat rhizosphere microbiome: evidences for a core microbiome across eight African and European soils. *FEMS Microbiology Ecology*, 96, 1–18.

Singer, D., Kosakyan, A., Seppey, C.V.W., Pillonel, A.M., Fernandez, L.D., Fontaneto, D. et al. (2018) Environmental filtering and phylogenetic clustering correlate with the distribution patterns of cryptic protist species. *Ecology*, 99, 904–914.

Sosa-Hernández, M.A., Roy, J., Hempel, S., Kautz, T., Köpke, U., Uksa, M. et al. (2018) Subsoil arbuscular mycorrhizal fungal communities in arable soil differ from those in topsoil. *Soil Biology and Biochemistry*, 117, 83–86.

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

Uksa, M., Fischer, D., Welzl, G., Kautz, T., Köpke, U. & Schloter, M. (2018) Community structure of prokaryotes and their functional potential in subsoils is more affected by spatial heterogeneity than by temporal variations. *Soil Biology and Biochemistry*, 75, 197–201.

von Humboldt, A. & Bonpland, A. (1814) *Voyage aux régions équinoxiales du nouveau monde*. Paris: F. Schoell.

Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261–5267.

Washburne, A.D., Silverman, J.D., Leff, J.W., Bennett, D.J., Darcy, J. L., Mukherjee, S. et al. (2017) Phylogenetic factorization of compositional data yields lineage-level associations in microbiome datasets. *PeerJ*, 5, e2969.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Roy, J., Mazel, F., Dumack, K., Bonkowski, M. & Rillig, M.C. (2022) Hierarchical phylogenetic community assembly of soil protists in a temperate agricultural field. *Environmental Microbiology*, 24(11), 5498–5508. Available from: https://doi.org/10.1111/1462-2920.16134