Sample size shapes metabarcoding-driven biodiversity assessments across body sizes in soil

Lu Wang [a b c]*, April Leonar [b c], Simone Cesarz [b c], Nico Eisenhauer [b c], Antonis Chatzinotas [a b c], Stephanie D. Jurburg [a b]

a Department of Applied Microbial Ecology, Helmholtz Centre for Environmental Research GmbH - UFZ, Leipzig, Germany

b German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

c Institute of Biology, Leipzig University, Leipzig, Germany

*Corresponding authors: lu.wang.env4all@gmail.com. Postal address: Permoserstraße 15, 04318 Leipzig, Germany

Abstract

Understanding how sample size influences biodiversity detection across taxonomic groups differing in body size is critical for designing robust and cost-efficient metabarcoding studies of soil eukaryotes. Using a soil mass gradient (0.25-32 g) combined with a universal 18S rRNA metabarcoding approach, we quantified how sample mass shapes diversity estimates across eukaryotic taxa. Diversity metrics (richness and inverse Simpson diversity; q = 0, 2) and community dispersion exhibited clear body size-dependent responses. Larger-bodied taxa (e.g., *Nematoda*, *Collembola*, *Insecta*) showed pronounced increases in detected diversity and reduced community dispersion with increasing soil mass, indicating that small soil samples fail to capture their full diversity. Conversely, microeukaryotic groups such as fungi and protists displayed weak or even negative relationships with increasing soil mass, implying limited improvement in detection with increased sampling effort. These findings demonstrate

that soil sample sizes are strongly modulated by organismal body size and spatial distribution. We propose that a taxon-specific analytical framework can enhance both the ecological representativeness and cost efficiency in metabarcoding-based soil biodiversity assessments and monitoring.

Introduction

Soil eukaryotes encompass a diverse array of organisms, including fungi, protists, nematodes, arthropods, and various other micro- to macro-faunal groups that collectively drive key ecosystem functions such as nutrient cycling, organic matter decomposition, and soil structure formation (Blouin et al., 2013; Eisenhauer et al., 2026; Geisen et al., 2018; Lee and Foster, 1991; Lehmann et al., 2017). Information about the diversity of soil biota provides critical insights into the complexity of ecosystem functions and trophic networks, and supports a mechanistic understanding of ecosystem resistance and resilience under environmental change (Phillips et al., 2024).

The advent of high-throughput DNA metabarcoding has opened new avenues to characterize these complex communities, offering insights into both their taxonomic and functional diversity across different spatial and temporal scales (Aslani et al., 2022; De Gruyter et al., 2020; Köninger et al., 2023; Seppey et al., 2017), and allowing for the simultaneous analysis of most groups of soil biota. Yet, methodological factors, especially aspects of sampling design, continue to shape the accuracy and representativeness of such biodiversity surveys (Jurburg et al., 2021; Lara et al., 2022; Liu et al., 2017). Among these, sample size (i.e., soil mass) is a critical determinant of detected diversity patterns, largely due to the heterogeneous and patchy distribution of different groups of soil biota (Bahram et al., 2016; Caruso et al., 2012; Ettema and Wardle, 2002; Liu et al., 2019; Quist et al., 2019). Smaller soil sample masses may fail to capture rare or spatially aggregated taxa, which can skew community profiles and lead to underestimates of overall biodiversity, while increasing compositional variability between samples (Dopheide et al., 2019; Kageyama and Toju, 2022; Nascimento et al., 2018; Wiesel et al., 2015).

Importantly, the effect of sample mass on the detectability of soil biota could be affected by the target organism’s body size, which is a phylogenetically conserved trait that reflects an

organism's abundance, life history, metabolic demands, and trophic role (Brown et al., 2004; Speakman, 2005; Woodward et al., 2005). By influencing resource use and energy flow, body size plays a central role in the structure and stability of food webs and ecosystems (Brown and Maurer, 1986; De Bie et al., 2012; Sheridan and Bickford, 2011; Woodward et al., 2005). Because larger-bodied eukaryotes tend to occur at lower densities and are more spatially patchy, increasing sample mass should disproportionately increase the likelihood of encountering these taxa. Consequently, alpha diversity is expected to rise more steeply with sample mass for larger-bodied groups than for smaller ones. Beyond effects on alpha diversity, spatial aggregation also influences variation in beta diversity, as smaller sample masses amplify among-replicate variability in community composition, or dispersion. We therefore expect increasing sample mass to reduce compositional dispersion most strongly for large-bodied taxa, whereas small-bodied taxa should remain comparatively heterogeneous. Nevertheless**,** despite these mechanistic expectations, the extent to which body size modulates sample-mass effects on both alpha diversity and compositional dispersion remains poorly understood in soil eukaryote metabarcoding studies. Clarifying this relationship is key to refining sampling strategies and ensuring that biodiversity assessments are both accurate and comparable across different studies and ecological contexts.

In this study, we employed a DNA metabarcoding approach to systematically explore how varying sample masses influence the detection and characterization of soil eukaryotic taxa across multiple body sizes. We hypothesized that (a) alpha diversity responses to soil mass increase with body size, (b) sample dispersion declines with soil mass, particularly in larger-bodied taxa. By examining taxonomic groups spanning the body size spectrum from 10 to 20000 μm, we aimed to uncover mass-dependent sampling effects and provide practical, empirically grounded guidance for designing future soil biodiversity surveys.

Methods and materials

Study site and sampling design

Sampling was conducted at the PhytOakmeter grassland field site (Bad Lauchstädt, Germany, 51˚23’30”N, 11˚52’49”E, 118 m a.s.l.), characterized by a Haplic Chernozem soil and a sub-continental, temperate climate. To isolate the technical effects of sample mass from broad-scale environmental heterogeneity (e.g., inter-tree variability), we established five independent sampling plots around a 2-m radius of a *Quercus robur*. From each plot, 3-5 5 cm soil cores were pooled, and thoroughly homogenized with a 2 mm sieve to create five composite soil replicates (~1 kg each). Subsequently, subsamples of each composite were weighed to create a soil mass gradient consisting of eight levels (0.25, 0.5, 1, 2, 4, 8, 16, and 32 g) for each of the five replicates, resulting in a total of 40 samples.

DNA extraction, PCR amplification, and Illumina sequencing

DNA was extracted from all 40 samples using FastDNA™ SPIN Kit for Soil, 50 mL tubes (MP Biomedicals, USA), according to the manufacturer's instructions. Importantly, to accommodate samples exceeding the kit’s capacity (max. 10 g), we implemented a split-and-pool strategy: for 16 g and 32 g samples, soil was split into 8 g aliquots (two and four aliquots, respectively), extracted in parallel, and the resulting lysates were pooled prior to downstream purification to ensure representativeness.

DNA extracts were quantified with Qubit dsDNA HS assay fluorometric system (Invitrogen, USA) and diluted to 5 ng/µL for PCR amplifications. The nuclear small subunit ribosomal RNA gene (18S rRNA gene) was chosen as the marker gene for eukaryotic metabarcoding and a universal 18S rRNA primer set (1391f/EukBr) from the Earth Microbiome Project was used to amplify the target V9 region (Amaral-Zettler et al., 2009). The 5′ ends of the primers were tagged with Illumina adapters required for multiplexing samples. The first PCR was performed in a reaction

mixture of 25 µL consisting of 12.5 µL MyTaq 2x mix buffer (Meridian Bioscience), 1 µL BSA (2.5 mM), 1 µL of each primer (25 µM), 3 µL DNA template and 6.5 µL PCR water. The PCR conditions consisted of an initial denaturation at 94 °C for 3 min, 35 cycles of denaturation at 94 °C for 45 s, annealing at 57 °C for 60 s, and elongation at 72 °C for 90 s, followed by a final elongation at 72 °C for 10 min. PCR amplicons were purified using AMPure XP beads and subsequently used as DNA templates in the index PCR step, following the manufacturer's suggested protocols. Index PCR amplicons were purified using AMPure XP beads, quantified with the Qubit fluorometric system, diluted to 4 nM, and subsequently pooled. Paired-end sequencing (2 × 300 bp) was performed on the Illumina MiSeq platform (Illumina Inc., San Diego, CA, USA).

Sequence processing and data analyses

Raw data processing was done with the dada2 pipeline (Callahan et al., 2016) in R software v 4.5.0 (R Core Team, 2025). Reads were truncated with *FilterAndTrim* function of dada2, with parameters set as trimLeft = 14, 24; truncLen= (165,155); maxN = 0; maxEE = (4, 4); truncQ = 2. Paired reads were merged using *mergePairs* function with minOverlap parameter as default. Taxonomy was assigned using PR2 database v 5.0.0 (Guillou et al., 2013) using the dada2 algorithm. Tables with the amplicon sequence variants (ASVs) were analyzed in R using the phyloseq (McMurdie and Holmes, 2013) and vegan (Oksanen et al., 2025) packages.

Missing ranks were labeled as "unidentified" in the taxonomy table. ASVs were grouped into taxonomic groups according to taxonomy and representative body size (Luan et al., 2020; Zinger et al., 2019). Fungi, algae, plants, bacteria, and archaea were grouped using domain- or class-level taxonomy. Metazoans were classified at the phylum level (e.g., *Nematoda*, *Annelida, Mollusca*), with Arthropoda and Annelida further classified by family when available. Within Annelida, *Enchytraeidae* were treated as intermediate-bodied taxa, whereas Lumbricidae and all other annelid taxa and unidentified taxa were grouped as large-bodied taxa ("Annelida:

Lumbricidae&others"). Protists were grouped based on eukaryotic supergroups and divisions (e.g., *Rhizaria*, *Amoebozoa*, *Alveolata*). Oomycetes (e.g., *Peronosporales*, *Saprolegniales*) were classified separately as *Oomycota*. Bacterial, archaeal, and plant taxa were excluded from downstream analyses, as their presence was not relevant to the primer target or ecological focus of this study.

To evaluate how soil sample size influences soil eukaryotic diversity across different body sizes, we measured Hill numbers (Chao et al., 2014; Jost, 2006) including observed richness (q = 0), exponential Shannon diversity (q = 1), inverse Simpson index (q = 2), and the dispersion of Bray-Curtis dissimilarities, using the *betadisper* function of the vegan package.

For each defined taxonomic group, taxa filtered to retain only those present in at least one sample. Rarefaction was performed within each group to the minimum sequencing depth among samples, and groups were excluded if the five lowest-depth samples had fewer than 100 reads. For alpha diversity, observed richness, exponential Shannon, and inverse Simpson indices were calculated using the microbiome package (Lahti and Shetty, 2017), and exponential Shannon diversity (q = 1) was obtained by exponentiating the Shannon index.

GLM families were selected based on the data distributions of the different diversity measurements. A Poisson generalized linear model (GLM with family = poisson) was used to model observed richness, whereas Gaussian GLMs (family = gaussian) were applied for Hill numbers with q = 1 and q = 2. For beta dispersion, a linear model (lm(Dispersion ~ log2(soil_g)) was fitted. Taxonomic groups were annotated with representative body size information (Luan et al., 2020; Zinger et al., 2019) and assigned to small-, intermediate-, or large-bodied groups. Group-level slope estimates were then analyzed using random-effects meta-regression using rma function in the metafor package (Viechtbauer, 2010), with body size as a moderator. This allowed evaluation of size-dependent responses to sample size across diversity metrics. Outputs included group-specific slope plots, meta-analysis forest plots, and summary tables.

To assess how similarity in community structure between taxonomic groups varied along a soil mass gradient, we performed pairwise Mantel tests (999 permutations) of Bray-Curtis dissimilarities within each soil mass level between taxonomic groups.

Saturation dynamics for Richness, exponential Shannon, and inverse Simpson were modeled using the Michaelis-Menten equation ($y = V_{max} \cdot m / (K_m + m)$) via Non-linear Least Squares (NLS). The saturation threshold (Sat_{95}) was calculated as $19 \cdot K_m$, corresponding to the soil mass required to reach 95% of the theoretical maximum (V_{max}).

Results

Across all taxonomic groups, richness (observed ASVs, q = 0) responded to soil sample mass, but this relationship varied across body size groups. For several intermediate- and large-bodied taxa (e.g., *Nematoda*, *Arachnida*, *Collembola*, *Insecta*), increasing sample mass led to a higher detected richness (Fig. 1A, left), while small taxa displayed weaker or negative responses. A meta-regression supported this pattern, indicating larger-bodied groups had stronger and more significant positive relationships between sample mass and richness than smaller-bodied taxa ($p < 0.01$; Fig. 1A, right). The exponential Shannon (q = 1; Fig. S1) and inverse Simpson (q = 2; Fig. 1B) indices showed similar body size-dependent trends. Larger-bodied groups showed consistently stronger positive responses to increasing soil mass, indicating that common and dominant taxa were more effectively detected when larger sample sizes were used. In contrast, for small-bodied taxa such as fungi and protists, diversity estimates tended to plateau or decline as soil mass increased.

In comparison, community dispersion decreased with increasing soil mass across all body size groups, with the exception of *Protura*, which showed high variability (Fig. 2). Notably, larger-bodied groups such as *Annelida (Lumbricidae)* exhibited the strongest negative relationships, indicating a substantial reduction in sample-to-sample variability with increased sample mass.

To identify how soil sample mass influences inter-group compositional relationships (i.e., interdependent ecological structures), we performed pairwise Mantel tests between taxonomic groups across all soil mass levels. Few significant correlations ($p \leqslant 0.05$) were detected (Table S1), occurring occasionally across soil masses and varying among taxa. Positive correlations were observed between certain fungal and arthropod groups, but no consistent pattern with soil mass was evident.

To determine the minimum soil mass required for biodiversity saturation across different body sizes, we modeled the accumulation dynamics of species richness, exponential Shannon, and inverse Simpson indices using Michaelis-Menten equations. The results revealed that the soil mass required for richness saturation was strongly dependent on organism body size (Fig. 3). Specifically, small-bodied taxa (e.g., Fungi and Protists) reached saturation at minimal soil masses (often < 0.25 g), whereas intermediate and large-bodied taxa required significantly higher masses to achieve saturation. This size-dependent pattern was consistent across diversity metrics, with exponential Shannon and inverse Simpson indices exhibiting similar accumulation dynamics (Table S2).

Discussion

Understanding how organismal traits shape biodiversity detection is essential for developing effective sampling strategies for soil ecosystems and monitoring. Here, we experimentally examined how sample size influences biodiversity estimates across taxonomic groups differing in body size.

Our study demonstrates a clear body size-related response of alpha diversity estimates to increasing soil sample mass across a wide range of taxonomic groups. Specifically, alpha diversity (q = 0, 1, and 2) increased more steeply with soil mass in larger-bodied taxa such as *Nematoda*, *Arachnida*, *Collembola*, and *Insecta*, indicating that larger-bodied organisms are

underrepresented in low soil sample masses. This pattern likely stems from the lower population density, greater spatial heterogeneity, and higher motility of larger soil biota (Bardgett and van der Putten, 2014; Decaëns, 2010; Ettema and Wardle, 2002), which together reduce detection probability at fine spatial scales, necessitating greater sampling effort (e.g. larger soil mass) to capture community composition effectively. In contrast, smaller-bodied taxa such as fungi and protists showed weak or even negative relationships between diversity and soil mass, with most relationships being statistically non-significant. This suggests that increasing sample mass does not substantially enhance the detectable alpha diversity estimates of microeukaryotic taxa. In addition to alpha diversity, community dispersion, i.e., the variance in composition among replicates, decreased consistently with increasing sample mass across all taxonomic groups. This suggests that larger samples improve representativeness by integrating across spatial and biological heterogeneity, thereby reducing stochastic variation (Decaëns, 2010; Ettema and Wardle, 2002; Zinger et al., 2019). Pairwise Mantel tests revealed no consistent pattern, rejecting the notion that compositional correlations between small and larger-bodied taxa increase with larger soil masses. We acknowledge that our experimental design (5 replicates per soil mass) likely reduced our ability to detect these subtle trends.

Previous research on soil biodiversity assembly supports our finding that larger sample sizes are needed to capture the full community composition of larger-bodied taxa, whose assembly is more strongly shaped by stochastic processes. Zinger et al. (2019) demonstrated that body size predicts assembly mechanisms in tropical soils, with mesofauna (e.g., *Arthropoda*, *Annelida*) exhibiting greater stochasticity and spatial heterogeneity than microbial taxa, such as bacteria and protists. Similarly, Aslani et al. (2022) reported that as body size increases, the influence of environmental selection weakens, whereas the contribution of ecological drift becomes more pronounced. Beyond organismal traits, methodological constraints, particularly sampling effort, sequencing coverage, and read depth can also amplify apparent stochasticity in community

assembly. Limited sampling or insufficient sequencing depth may underrepresent rare or spatially dispersed taxa, thereby enhancing the observed randomness of community patterns. Together, these biological and methodological factors likely explain why larger-bodied taxa exhibit a stronger positive response of alpha diversity to increasing soil mass.

In molecular analyses of soil microbial communities, researchers commonly utilize relatively small sample quantities, often between 0.25 to 1 g, since previous evidence has demonstrated that such amounts can yield representative and reproducible results. For fungal communities, even larger soil amounts, generally exceeding 1 g, are recommended to apply to adequately capture their more heterogeneous spatial distribution and community diversity (Kageyama and Toju, 2022; Li et al., 2023; Penton et al., 2016; Ranjard et al., 2003). In contrast, molecular eDNA studies targeting micro- to mesofaunal assemblages, such as nematodes and collembolans, remain less standardized with respect to soil input. Some studies have used small quantities comparable to those applied in microbiome studies (Gong et al., 2023; Seppey et al., 2017), often improving representativeness by pooling several replicates (Hermans et al., 2022; Köninger et al., 2023). Others, however, have employed larger sample sizes, up to 10 g, to obtain a more comprehensive representation of soil community diversity (Donhauser et al., 2023; Gong et al., 2024).

This variability reflects ongoing methodological uncertainty regarding the optimal sampling size, which partly arises from trade-offs inherent in molecular procedures. Larger soil quantities can improve representativeness but may also increase the concentration of PCR inhibitors such as humic acids and metals, reducing amplification efficiency (Schrader et al., 2012; Zinger et al., 2016). Moreover, working with large soil masses can raise both extraction difficulty and processing costs, making it an inefficient strategy for routine or long-term biodiversity surveys. For taxa with rigid or resistant body structures, such as chitinous or sclerotized fauna, cell walls and cuticles can hinder complete lysis during extraction, resulting in low DNA release efficiency

(Martoni et al., 2022; Waeyenberge et al., 2019). This intrinsic physiological limitation constrains DNA recovery even when soil mass is increased, underscoring the need to improve extraction protocols rather than simply maximizing sample size. Therefore, an optimal soil mass should balance ecological representativeness, extraction efficiency, and technical feasibility.

Our findings align with earlier evidence linking sample size to biodiversity estimates, and provide guidance for future studies. Dopheide et al. (2019) reported higher arthropod diversity with larger soil volumes, whereas DNA extraction size had a minimal effect on prokaryote and microeukaryote diversity. Likewise, Kageyama and Toju (2022) found that bacterial and fungal diversity metrics were largely independent of soil mass, while nematode diversity increased significantly with sample size, recommending over 20 g of soil for comprehensive cross-taxa diversity analyses. Our study extends these findings by applying a consistent DNA metabarcoding workflow across a continuous range of soil masses (0.25-32 g) and body-size categories. Our primer selection reflects a generalist approach that is increasingly common in soil biology (Jurburg et al., 2021), and the consistent analytical framework used allows robust comparison among groups, especially in the context of large-scale monitoring frameworks. We acknowledge that this study was conducted in a single soil type. This design was intentionally chosen to minimize environmental heterogeneity, allowing us to clearly isolate the technical impact of sample mass. While future work is needed to validate specific mass thresholds across diverse biomes, the body-size-dependent detection bias observed here highlights a fundamental methodological challenge that likely affects soil monitoring globally.

Our non-linear saturation model (Michaelis-Menten) demonstrates that the sample mass required to capture species richness is strongly dependent on organism body size. While micro-eukaryotes (e.g., fungi and protists) reach saturation rapidly with minimal input, larger taxa exhibit more gradual saturation profiles, with Nematoda stabilizing at approximately 2-3 g and larger groups (e.g., Arthropoda, Annelida) often requiring 3-5 g. This distinct taxon-specific

sensitivity indicates that standard low-input protocols (e.g., 0.25 g) systematically underestimate the richness of intermediate- and large-bodied taxa. These findings highlight that in studies targeting specific soil taxa, the choice of sample should be tailored to the body size and ecological characteristics of the target groups. Conversely, in more general studies targeting all soil biota (e.g., employing a universal 18S rRNA marker region), we recommend a stratified analytical approach that evaluates taxa separately based on their body size/specific saturation thresholds rather than applying a universal analytical standard (Jurburg et al., 2021).

Acknowledgements

This research was conducted as part of the PhytOakmeter consortium, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 507084794, and the Swiss National Science Foundation (SNSF). We thank the Helmholtz Centre for Environmental Research – UFZ for logistic support and permission to do research on their sites. We would like to especially acknowledge local support from Experimental field station Bad Lauchstädt (Dr. Ines Merbach). We thank Nicole Steinbach for technical support and Microbial Interaction Ecology Group, Helmholtz Centre for Environmental Research GmbH – UFZ for providing research facilities. We thank Dr. Ulrike Schlägel for valuable statistical consultation at UFZ. To ensure data FAIRness, we utilized the DataPLANT personal assistance network and its tool stack, centered around the DataHUB.

CRediT authorship contribution statement

Lu Wang: Writing – original draft, Writing – review & editing, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Investigation. April Lyn Leonar: Writing – review & editing. Simone Cesarz: Writing – review & editing, Supervision, Resources. Nico Eisenhauer: Writing – review & editing, Supervision, Resources. Antonis Chatzinotas: Writing – review & editing, Supervision, Resources. Stephanie Jurburg: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Data Accessibility

Sequencing data generated in this study are available in the NCBI Sequence Read Archive under BioProject accession number PRJNA1359375. Downstream analytical workflows are accessible via GitLab (https://git.nfdi4plants.org/lu.wang.env4all/euk_metabarcoding_body_size).

References

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. https://doi.org/10.1371/journal.pone.0006372

Aslani, F., Geisen, S., Ning, D., Tedersoo, L., Bahram, M., 2022. Towards revealing the global diversity and community assembly of soil eukaryotes. Ecol. Lett. 25, 65–76. https://doi.org/10.1111/ele.13904

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. ISME J. 10, 885–896. https://doi.org/10.1038/ismej.2015.164

Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. Nature 515, 505–511. https://doi.org/10.1038/nature13855

Blouin, M., Hodson, M.E., Delgado, E.A., Baker, G., Brussaard, L., Butt, K.R., Dai, J., Dendooven, L., Peres, G., Tondoh, J.E., Cluzeau, D., Brun, J.-J., 2013. A review of earthworm impact on soil function and ecosystem services. Eur. J. Soil Sci. 64, 161–182. https://doi.org/10.1111/ejss.12025

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., West, G.B., 2004. Toward a Metabolic Theory of Ecology. Ecology 85, 1771–1789. https://doi.org/10.1890/03-9000

Brown, J.H., Maurer, B.A., 1986. Body size, ecological dominance and Cope's rule. Nature 324, 248–250. https://doi.org/10.1038/324248a0

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869

Caruso, T., Taormina, M., Migliorini, M., 2012. Relative role of deterministic and stochastic determinants of soil animal community: a spatially explicit analysis of oribatid mites. J. Anim. Ecol. 81, 214–221. https://doi.org/10.1111/j.1365-2656.2011.01886.x

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., Ellison, A.M., 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecol. Monogr. 84, 45–67. https://doi.org/10.1890/13-0133.1

De Bie, T., De Meester, L., Brendonck, L., Martens, K., Goddeeris, B., Ercken, D., Hampel, H., Denys, L., Vanhecke, L., Van der Gucht, K., Van Wichelen, J., Vyverman, W., Declerck, S. a. J., 2012. Body size and dispersal mode as key traits determining metacommunity structure of aquatic organisms. Ecol. Lett. 15, 740–747. https://doi.org/10.1111/j.1461-0248.2012.01794.x

Decaëns, T., 2010. Macroecological patterns in soil communities. Glob. Ecol. Biogeogr. 19, 287–302. https://doi.org/10.1111/j.1466-8238.2009.00517.x

De Gruyter, J., Weedon, J.T., Bazot, S., Dauwe, S., Fernandez-Garberí, P.-R., Geisen, S., De La Motte, L.G., Heinesch, B., Janssens, I.A., Leblans, N., Manise, T., Ogaya, R., Löfvenius, M.O., Peñuelas, J., Sigurdsson, B.D., Vincent, G., Verbruggen, E., 2020. Patterns of local, intercontinental and interseasonal variation of soil bacterial and eukaryotic microbial communities. FEMS Microbiol. Ecol. 96, fiaa018. https://doi.org/10.1093/femsec/fiaa018

Donhauser, J., Briones, M.J.I., Mikola, J., Jones, D.L., Eder, R., Filser, J., Frossard, A., Krogh, P.H., Sousa, J.P., Cortet, J., Desie, E., Domene, X., Djuric, S., Hackenberger, D., Jimenez, J.J., Iamandei, M., Rissmann, C., Schmidt, O., Shanskiy, M., Silfver, T., Vancampenhout, K., Vasutova, M., Velizarova, E., Frey, B., 2023. Extracting DNA from soil or directly from isolated nematodes indicate dissimilar community structure for Europe-wide forest soils. Soil Biol. Biochem. 185, 109154. https://doi.org/10.1016/j.soilbio.2023.109154

Dopheide, A., Xie, D., Buckley, T.R., Drummond, A.J., Newcomb, R.D., 2019. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods Ecol. Evol. 10, 120–133. https://doi.org/10.1111/2041-210X.13086

Eisenhauer, N., Sünnemann, M., Pollierer, M.M., Sun, X., Bardgett, R.D., Bartkowski, B., Delgado-Baquerizo, M., Dirilgen, T., Guerra, C.A., Mathieu, J., Niklaus, P.A., Romero, F., Ristok, C., Seeber, J., Steinwandter, M., Stewart, J., van der Heijden, M., van der Putten, W., Potapov, A., 2026. Soil biodiversity effects on ecosystems. Nat. Rev. Biodivers. 2, 76–91. https://doi.org/10.1038/s44358-025-00123-z

Ettema, C.H., Wardle, D.A., 2002. Spatial soil ecology. Trends Ecol. Evol. 17, 177–183. https://doi.org/10.1016/S0169-5347(02)02496-5

Geisen, S., Mitchell, E.A.D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L.D., Jousset, A., Krashevska, V., Singer, D., Spiegel, F.W., Walochnik, J., Lara, E., 2018. Soil protists: a fertile frontier in soil biology research. FEMS Microbiol. Rev. 42, 293–323. https://doi.org/10.1093/femsre/fuy006

Gong, X., Qiao, Z., Yao, H., Zhao, D., Eisenhauer, N., Scheu, S., Liang, C., Liu, M., Zhu, Y.-G., Sun, X., 2024. Urbanization simplifies soil nematode communities and coincides with decreased ecosystem stability. Soil Biol. Biochem. 190, 109297. https://doi.org/10.1016/j.soilbio.2023.109297

Gong, X., Sun, X., Thakur, M.P., Qiao, Z., Yao, H., Liu, M., Scheu, S., Zhu, Y.-G., 2023. Climate and edaphic factors drive soil nematode diversity and community composition in urban ecosystems. Soil Biol. Biochem. 180, 109010. https://doi.org/10.1016/j.soilbio.2023.109010

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J.R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W.H.C.F., Lara, E., Le Bescot, N., Logares, R., Mahé, F., Massana, R., Montresor, M., Morard, R., Not, F., Pawlowski, J., Probert, I., Sauvadet, A.-L., Siano, R., Stoeck, T., Vaulot, D., Zimmermann, P., Christen, R., 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41, D597–D604. https://doi.org/10.1093/nar/gks1160

Hermans, S.M., Lear, G., Buckley, T.R., Buckley, H.L., 2022. Environmental DNA sampling detects between-habitat variation in soil arthropod communities, but is a poor indicator of fine-scale spatial and seasonal variation. Ecol. Indic. 140, 109040. https://doi.org/10.1016/j.ecolind.2022.109040

Jost, L., 2006. Entropy and diversity. Oikos 113, 363–375. https://doi.org/10.1111/j.2006.0030-1299.14714.x

Jurburg, S.D., Keil, P., Singh, B.K., Chase, J.M., 2021. All together now: Limitations and recommendations for the simultaneous analysis of all eukaryotic soil sequences. Mol. Ecol. Resour. 21, 1759–1771. https://doi.org/10.1111/1755-0998.13401

Kageyama, T., Toju, H., 2022. Effects of source sample amount on biodiversity surveys of bacteria, fungi, and nematodes in soil ecosystems. Front. Ecol. Evol. 10. https://doi.org/10.3389/fevo.2022.959945

Köninger, J., Ballabio, C., Panagos, P., Jones, A., Schmid, M.W., Orgiazzi, A., Briones, M.J.I., 2023. Ecosystem type drives soil eukaryotic diversity and composition in Europe. Glob. Change Biol. 29, 5706–5719. https://doi.org/10.1111/gcb.16871

Lahti, L., Shetty, S., 2017. Tools for microbiome analysis in R.

Lara, E., Singer, D., Geisen, S., 2022. Discrepancies between prokaryotes and eukaryotes need to be considered in soil DNA-based studies. Environ. Microbiol. 24, 3829–3839. https://doi.org/10.1111/1462-2920.16019

Lee, K.E., Foster, R.C., 1991. Soil fauna and soil structure. Soil Res. 29, 745–775. https://doi.org/10.1071/sr9910745

Lehmann, A., Zheng, W., Rillig, M.C., 2017. Soil biota contributions to soil aggregation. Nat. Ecol. Evol. 1, 1828–1835. https://doi.org/10.1038/s41559-017-0344-y

Li, T., Zhang, S., Hu, J., Hou, H., Li, K., Fan, Q., Wang, F., Li, L., Cui, X., Liu, D., Che, R., 2023. Soil sample sizes for DNA extraction substantially affect the examination of microbial diversity and co-occurrence patterns but not abundance. Soil Biol. Biochem. 177, 108902. https://doi.org/10.1016/j.soilbio.2022.108902

Liu, L., Liu, M., Wilkinson, D.M., Chen, H., Yu, X., Yang, J., 2017. DNA metabarcoding reveals that 200-μm-size-fractionated filtering is unable to discriminate between planktonic microbial and large eukaryotes. Mol. Ecol. Resour. 17, 991–1002. https://doi.org/10.1111/1755-0998.12652

Liu, T., Hu, F., Li, H., 2019. Spatial ecology of soil nematodes: Perspectives from global to micro scales. Soil Biol. Biochem. 137, 107565. https://doi.org/10.1016/j.soilbio.2019.107565

Luan, L., Jiang, Y., Cheng, M., Dini-Andreote, F., Sui, Y., Xu, Q., Geisen, S., Sun, B., 2020. Organism body size structures the soil microbial and nematode community assembly at a continental and global scale. Nat. Commun. 11, 6406. https://doi.org/10.1038/s41467-020-20271-4

Martoni, F., Piper, A.M., Rodoni, B.C., Blacket, M.J., 2022. Disentangling bias for non-destructive insect metabarcoding. PeerJ 10, e12981. https://doi.org/10.7717/peerj.12981

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Nascimento, F.J.A., Lallias, D., Bik, H.M., Creer, S., 2018. Sample size effects on the assessment of eukaryotic diversity and community structure in aquatic sediments using high-throughput sequencing. Sci. Rep. 8, 11737. https://doi.org/10.1038/s41598-018-30179-1

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H.B.A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M.O., Lahti, L., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C.J.F., Weedon, J., 2025. vegan: Community ecology package (manual).

Penton, C.R., Gupta, V.V.S.R., Yu, J., Tiedje, J.M., 2016. Size Matters: Assessing Optimum Soil Sample Size for Fungal and Bacterial Community Structure Analyses Using High Throughput Sequencing of rRNA Gene Amplicons. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00824

Phillips, H.R.P., Cameron, E.K., Eisenhauer, N., Burton, V.J., Ferlian, O., Jin, Y., Kanabar, S., Malladi, S., Murphy, R.E., Peter, A., Petrocelli, I., Ristok, C., Tyndall, K., Putten, W. van der, Beaumelle, L., 2024. Global changes and their environmental stressors have a significant impact on soil biodiversity—A meta-analysis. iScience 27. https://doi.org/10.1016/j.isci.2024.110540

Quist, C.W., Gort, G., Mooijman, P., Brus, D.J., van den Elsen, S., Kostenko, O., Vervoort, M., Bakker, J., van der Putten, W.H., Helder, J., 2019. Spatial distribution of soil nematodes

relates to soil organic matter and life strategy. Soil Biol. Biochem. 136, 107542. https://doi.org/10.1016/j.soilbio.2019.107542

R Core Team, 2025. R: a language and environment for statistical computing (manual). R Foundation for Statistical Computing, Vienna, Austria.

Ranjard, L., Lejon, D.P.H., Mougel, C., Schehrer, L., Merdinoglu, D., Chaussod, R., 2003. Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities. Environ. Microbiol. 5, 1111–1120. https://doi.org/10.1046/j.1462-2920.2003.00521.x

Schrader, C., Schielke, A., Ellerbroek, L., Johne, R., 2012. PCR inhibitors – occurrence, properties and removal. J. Appl. Microbiol. 113, 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Seppey, C.V.W., Singer, D., Dumack, K., Fournier, B., Belbahri, L., Mitchell, E.A.D., Lara, E., 2017. Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling. Soil Biol. Biochem. 112, 68–76. https://doi.org/10.1016/j.soilbio.2017.05.002

Sheridan, J.A., Bickford, D., 2011. Shrinking body size as an ecological response to climate change. Nat. Clim. Change 1, 401–406. https://doi.org/10.1038/nclimate1259

Speakman, J.R., 2005. Body size, energy metabolism and lifespan. J. Exp. Biol. 208, 1717–1730. https://doi.org/10.1242/jeb.01556

Viechtbauer, W., 2010. Conducting Meta-Analyses in *R* with the **metafor** Package. J. Stat. Softw. 36. https://doi.org/10.18637/jss.v036.i03

Waeyenberge, L., de Sutter, N., Viaene, N., Haegeman, A., 2019. New Insights Into Nematode DNA-metabarcoding as Revealed by the Characterization of Artificial and Spiked Nematode Communities. Diversity 11, 52. https://doi.org/10.3390/d11040052

Wiesel, L., Daniell, T.J., King, D., Neilson, R., 2015. Determination of the optimal soil sample size to accurately characterise nematode communities in soil. Soil Biol. Biochem. 80, 89–91. https://doi.org/10.1016/j.soilbio.2014.09.026

Woodward, G., Ebenman, B., Emmerson, M., Montoya, J.M., Olesen, J.M., Valido, A., Warren, P.H., 2005. Body size in ecological networks. Trends Ecol. Evol. 20, 402–409. https://doi.org/10.1016/j.tree.2005.04.005

Zinger, L., Chave, J., Coissac, E., Iribar, A., Louisanna, E., Manzi, S., Schilling, V., Schimann, H., Sommeria-Klein, G., Taberlet, P., 2016. Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. Soil Biol. Biochem. 96, 16–19. https://doi.org/10.1016/j.soilbio.2016.01.008

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., Gaucher, P., Gielly, L., Giguet-Covex, C., Iribar, A., Réjou-Méchain, M., Rayé, G., Rioux, D., Schilling, V., Tymen, B., Viers, J., Zouiten, C., Thuiller, W., Coissac, E., Chave, J., 2019. Body size determines soil community assembly in a tropical forest. Mol. Ecol. 28, 528–543. https://doi.org/10.1111/mec.14919

Figures

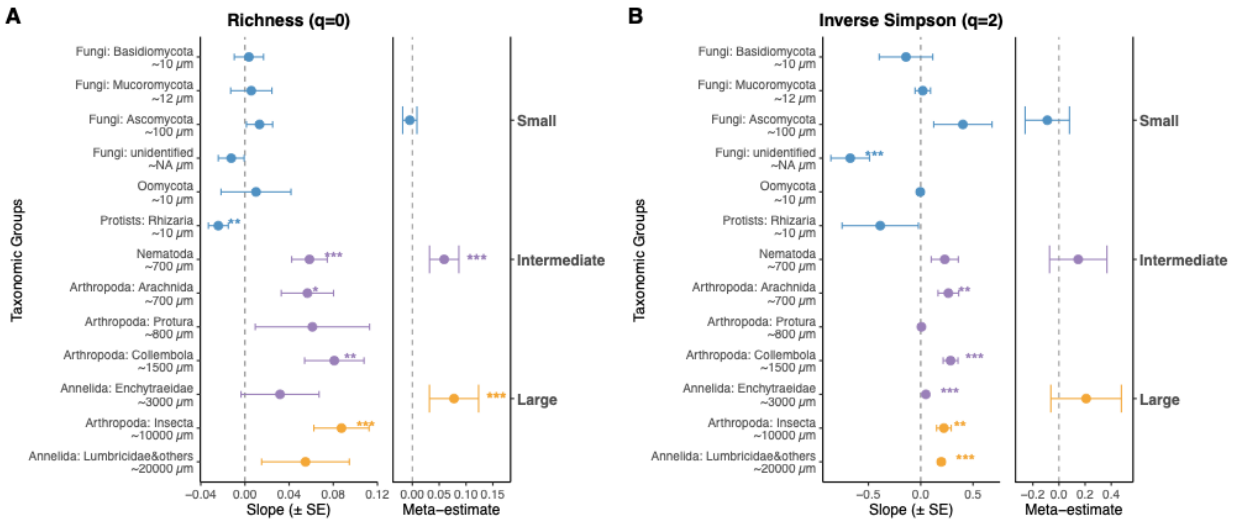

Fig 1. Soil sample mass effects on alpha diversity estimates across taxonomic groups categorized by body size. (A) Richness (observed ASVs, q = 0) and (B) Inverse Simpson diversity (q = 2) were modeled using GLMs with log-transformed soil mass as predictor. Each point in the main panels represents a taxonomic group; error bars indicate standard errors; asterisks indicate group-level significance of the soil mass slope (* p < 0.05, ** p < 0.01, *** p < 0.001). Groups are ordered by average body size rank from small to large. Meta-regression results (right panels) summarize effect sizes across body-size categories (small-, intermediate-, and large-bodied taxa). Asterisks here indicate whether the estimated slope for each size group is significantly different from zero.

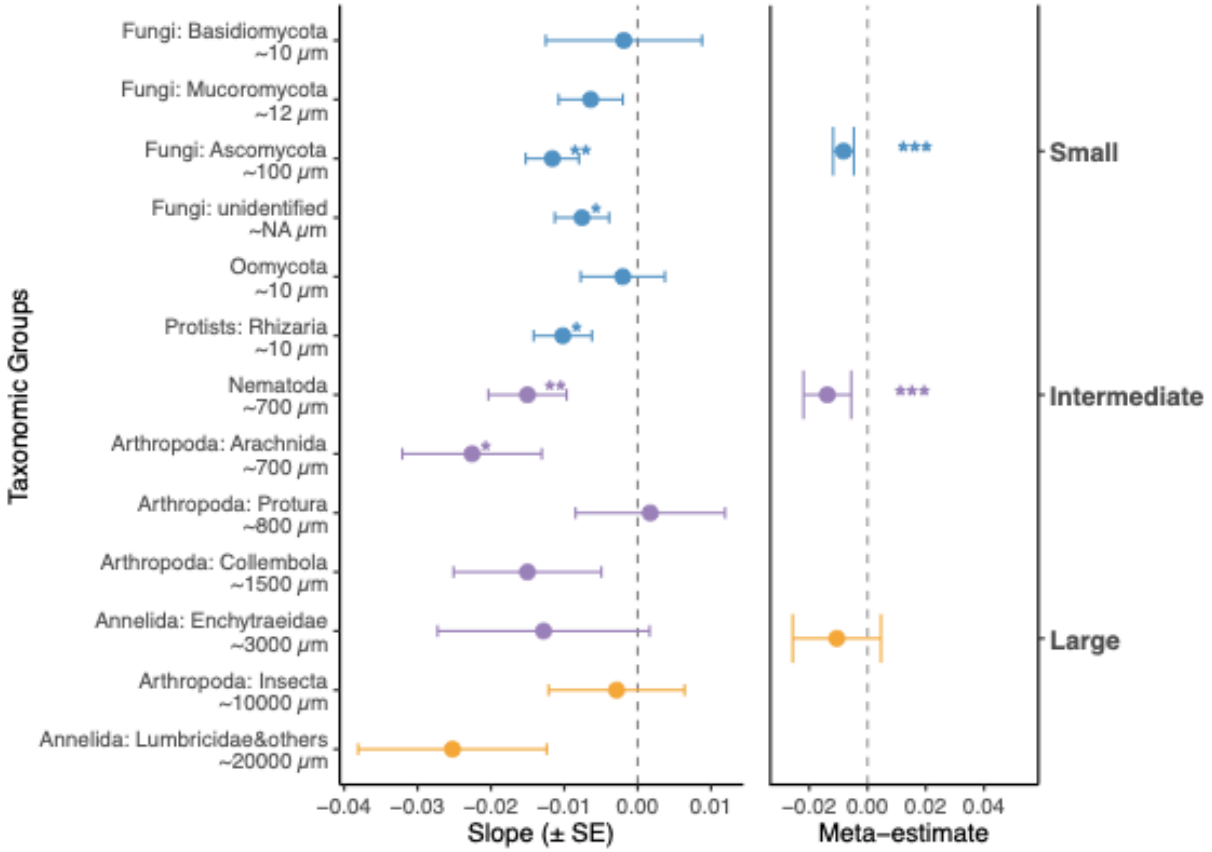

Fig 2. Soil sample mass effects on community dispersion (Bray-Curtis distance to group centroids) across taxonomic groups categorized by body size. Each point in the main panels represents a taxonomic group; error bars indicate standard errors; asterisks indicate group-level significance of the soil mass slope (* p < 0.05, ** p < 0.01, *** p < 0.001). Groups are ordered by average body size rank from small to large. Meta-

regression results (right panels) summarize effect sizes across body-size categories (small-, intermediate-, and large-bodied taxa). Asterisks here indicate whether the estimated slope for each size group is significantly different from zero.

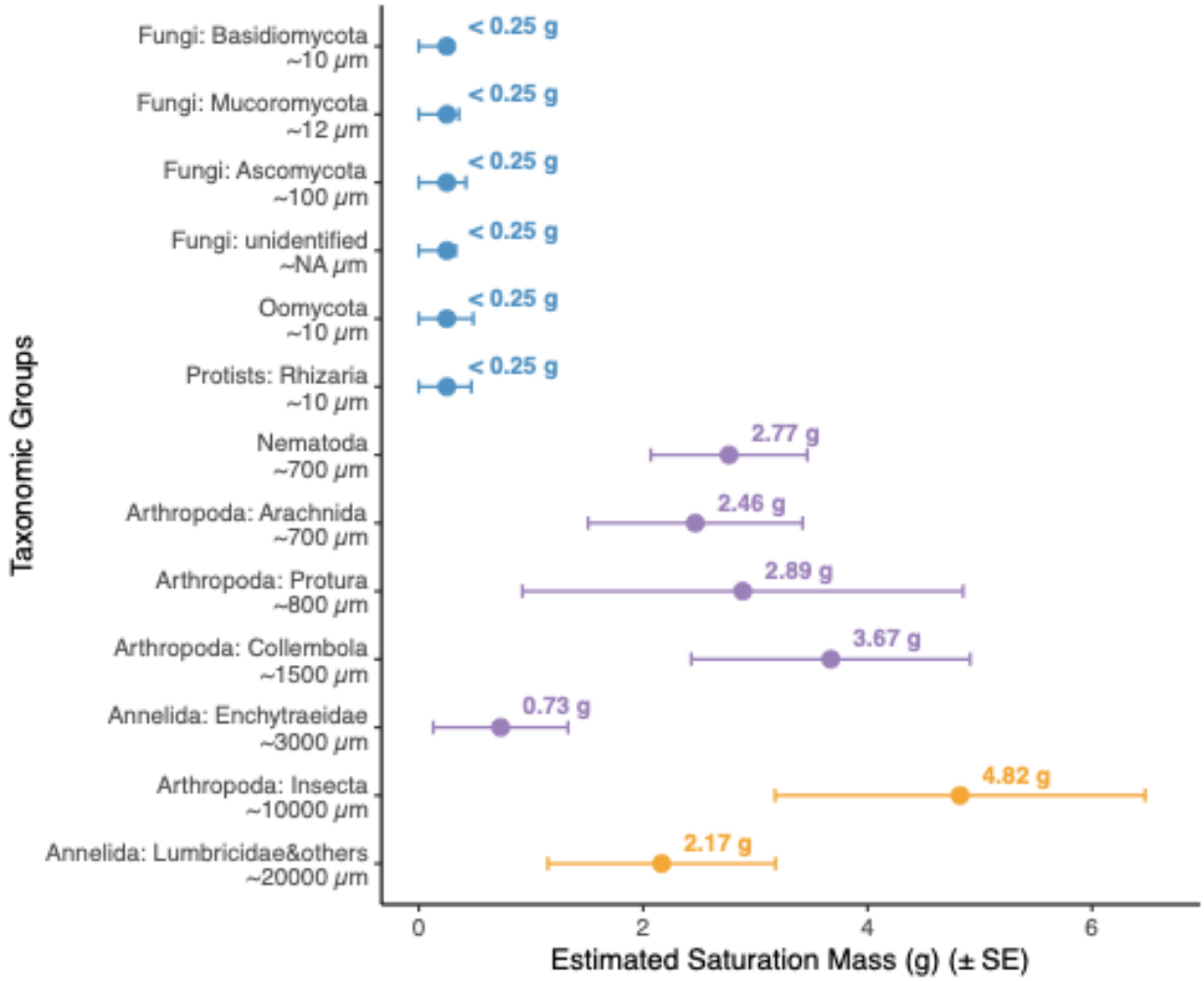

Fig 3. Minimum soil mass requirements for species richness saturation (Sat_{95}) across taxonomic groups categorized by body size. Each point represents a taxonomic group; error bars indicate standard errors (SE) of the estimated saturation point derived from Michaelis-Menten models. Groups are ordered by average body size from small to large. Text labels display the specific estimated saturation mass. Labels of "< 0.25 g" indicate saturation occurred at or below the minimum experimental sampling unit. Colors differentiate body size categories.

Supporting Information

Tables and Figures

Table S1. Significant Mantel correlations among taxonomic groups across soil mass gradients. Pairwise Mantel tests were performed between community dissimilarity matrices of all taxonomic groups within each soil mass level. Only correlations with $p \leqslant 0.05$ are shown. Each row represents a group pair that exhibited a statistically significant structural correlation (Mantel r) at a given soil mass.

Base_Group	Compare_Group	Soil_Mass	Mantel_r	P_value	N
Annelida_Enchytraeidae	Nematoda	1	0.66	0.050	5
Annelida_Enchytraeidae	Oomycota	32	0.76	0.042	5
Annelida_Enchytraeidae	Protists_Rhizaria	0.5	0.71	0.025	5
Annelida_Lumbricidae&others	Arthropoda_Insecta	0.25	0.56	0.008	5
Fungi_Basidiomycota	Annelida_Lumbricidae&others	8	0.77	0.017	5
Annelida_Lumbricidae&others	Fungi_Mucoromycota	16	0.57	0.017	5
Arthropoda_Arachnida	Arthropoda_Collembola	2	0.56	0.008	5
Arthropoda_Protura	Arthropoda_Arachnida	8	0.91	0.033	5
Fungi_Ascomycota	Arthropoda_Collembola	1	0.51	0.033	5
Fungi_Basidiomycota	Arthropoda_Collembola	0.25	0.90	0.017	5
Nematoda	Arthropoda_Insecta	4	0.67	0.033	5
Oomycota	Arthropoda_Insecta	2	0.48	0.042	5
Fungi_Basidiomycota	Arthropoda_Protura	0.5	0.86	0.050	5
Fungi_Basidiomycota	Arthropoda_Protura	2	0.37	0.033	5
Fungi_Basidiomycota	Arthropoda_Protura	32	0.91	0.017	5
Arthropoda_Protura	Oomycota	0.5	0.49	0.050	5
Fungi_Mucoromycota	Fungi_Ascomycota	0.25	0.68	0.050	5
Fungi_Basidiomycota	Oomycota	0.5	0.51	0.008	5
Fungi_Basidiomycota	Oomycota	1	0.77	0.033	5
Fungi_Basidiomycota	Oomycota	4	0.47	0.008	5
Oomycota	Fungi_Mucoromycota	1	0.58	0.050	5
Oomycota	Protists_Rhizaria	2	0.72	0.025	5
Oomycota	Protists_Rhizaria	8	0.74	0.033	5

Table S2. Minimum soil sampling mass required for biodiversity saturation (Sat_{95}) across body size groups. Saturation points were estimated using Michaelis-Menten models based on rarefied amplicon data. Data represent the estimated saturation mass (g), with R^2 in parentheses. Sat. (< 0.25): Saturation occurred at or below the minimum experimental sampling unit (indicating rapid stabilization). Proposed minimum composite sampling mass to ensure adequate coverage of the respective taxon.

Group	Richness	Shannon	Simpson	Mass_Suggest
Fungi_Basidiomycota	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	< 0.25 g
Fungi_Mucoromycota	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	< 0.25 g
Fungi_Ascomycota	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	0.34 (<.01)	~ 1 g
Fungi_unidentified	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	< 0.25 g
Oomycota	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	< 0.25 g
Protists_Rhizaria	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	Sat.(<0.25) (<.01)	< 0.25 g
Nematoda	2.77 (0.41)	2.21 (0.25)	1.83 (0.13)	~ 3 g
Arthropoda_Arachnida	2.46 (0.24)	4.89 (0.22)	3.71 (0.12)	~ 5 g
Arthropoda_Protura	2.89 (0.10)	1.20 (0.06)	0.87 (0.04)	~ 3 g
Arthropoda_Collembola	3.67 (0.31)	4.07 (0.37)	3.94 (0.31)	~ 5 g
Annelida_Enchytraeidae	0.73 (0.04)	1.93 (0.35)	1.16 (0.25)	~ 2 g
Arthropoda_Insecta	4.82 (0.32)	3.76 (0.19)	2.88 (0.13)	~ 5 g
Annelida_Lumbricidae&others	2.17 (0.15)	4.75 (0.34)	5.15 (0.36)	~ 6 g

527

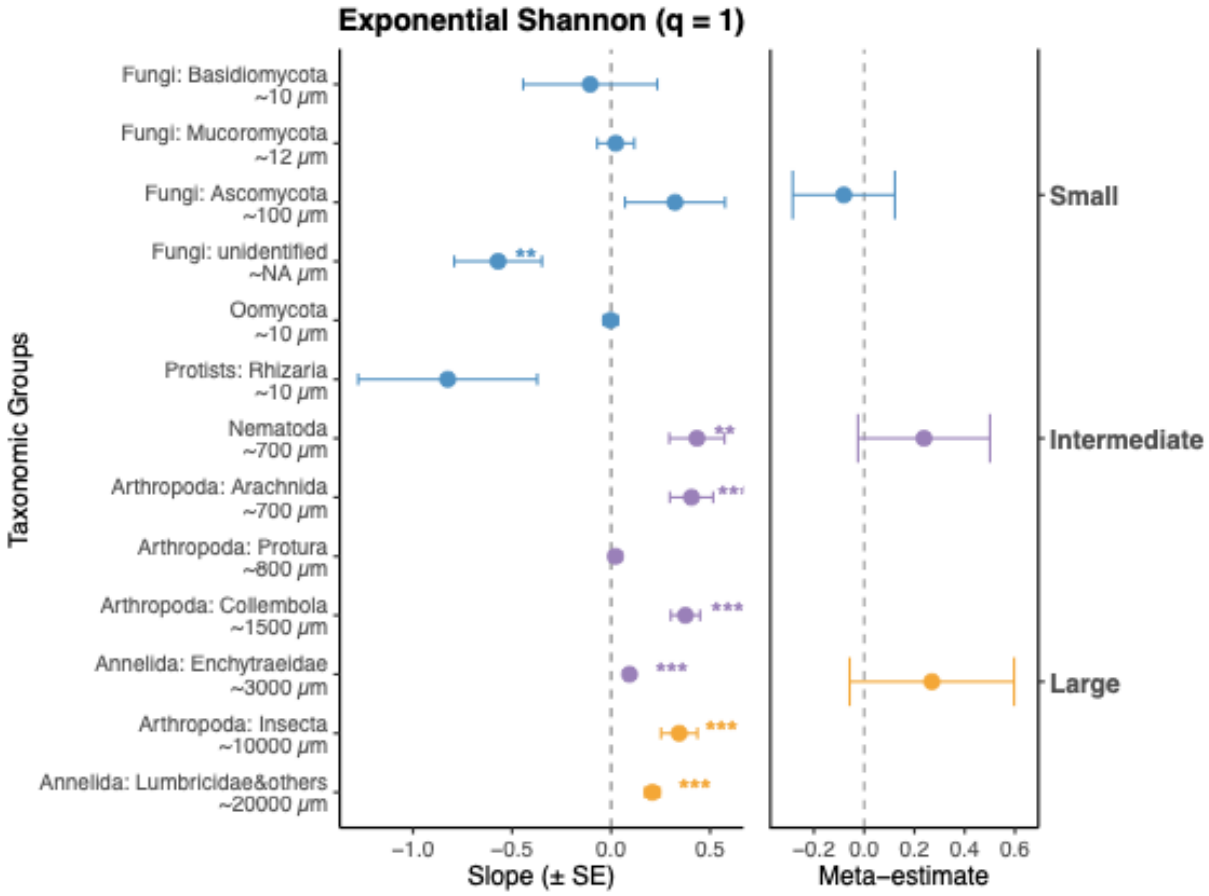

Fig. S2. Soil sample mass effects on Exponential Shannon diversity (q = 1) across taxonomic groups categorized by body size. Exponential Shannon diversity was modeled using generalized linear models (GLMs) with log-transformed soil mass as the predictor variable. Each point in the main panels represents a taxonomic group; error bars indicate standard errors; asterisks denote significance levels for the soil mass slope compared to zero (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Groups are ordered by average body size rank from small to large. Meta-regression results (right panels) summarize effect sizes across body-size categories (small-, intermediate-, and large-bodied taxa), with asterisks indicating whether each size-group slope differs significantly from zero.