Report on spatial sampling designs for long term biodiversity monitoring of soil biota

Stephanie D. Jurburg[1,2*], Maria Kostakou[1], Lu Wang[1,2], Antonis Chatzinotas [1,2, 3]

[1] Department of Applied Microbial Ecology, Helmholtz Centre for Environmental Research - UFZ, Leipzig, 04318, Germany

[2] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, 04103, Germany

[3]Institute of Biology, Leipzig University, Leipzig, 04103 Germany

*S.d.jurburg@gmail.com

Soil biota are extremely diverse and spatially heterogeneous. Sampling strategies that target the soil biota through metabarcoding must incorporate existing ecological theory and explicitly consider technical constraints. This report synthesizes empirical research from large-scale projects (SpaceMic, PhytOakMeter) to evaluate how spatial sampling design—specifically sample extent, number, and size—influences metabarcoding-derived biodiversity estimates across a range of soil biota. We find that the practice of homogenizing several soil samples per plot into one composite sample results in linear increases in plot diversity with sample number (Spearman's $\rho > 0.96$, $p < 0.001$), likely at the cost of sample completeness. Conversely, we found no effect of sample extent on plot-level diversity assessments, suggesting flexibility in plot sizing for monitoring programs but highlighting the importance of consistent plot sizing across a monitoring program. Finally, DNA extraction protocols, and their associated sample size, critically affect detectability: larger soil samples (10g vs. 0.5g) improved nematode diversity recovery, highlighting protocol-dependent biases. This study advocates for standardized, spatially explicit designs that take both a single and a composite sample per plot to ensure comparability in long-term datasets. Our insights provide actionable guidelines for enhancing the reproducibility and ecological relevance of soil metabarcoding in long-term monitoring initiatives.

Keywords: soil microbiome, metabarcoding, sampling design, biodiversity monitoring.

Background

The central role of spatial patterns in shaping ecological relationships is widely recognized for all organisms, and some of ecology's most fundamental theories (e.g., the distance-decay and taxa-area relationships, see below) relate to the placement of organisms in space. These theories have implications for the design of sampling schemes that ensure that the resulting data represent biological 'truth' to the greatest extent possible. Nevertheless, sampling designs in soil microbial ecology routinely ignore these guidelines, and a recent analysis found that only 10% of publications in soil microbial ecology mentioned the soil habitat, compared to ~30% across macro-ecology (Nunan 2017). Many factors contribute this phenomenon, including practical constraints imposed by the sequencing techniques applied to the samples (e.g., metabarcoding, shotgun metagenomics); the incompatibility of the small scales over which soil microbial ecology manifests (Ranjard *et al.* 2013), which is often incompatible with field sampling, especially when other organisms (e.g., plants) are also considered; and the relatively slow application of ecological principles to microbial ecology (Prosser *et al.* 2007).

Soil ecosystems sustain terrestrial life. Soil microbial communities are extremely heterogeneous and intricately linked to their environment (e.g., pH (Fierer & Jackson 2006), and despite the high diversity found in these communities within small samples (i.e., < 1g), available research suggests that microbial interactions occur over tens of micrometers (Gantner *et al.* 2006), and that communities of interacting microbes in soil are small, localized, and species-poor (Nunan 2017). It is therefore important that sampling designs consider both the immediate environment of the sample, and that they attempt to approximate the scale of the interacting community as much as possible.

The growing recognition of the importance of soil microbial diversity (Averill *et al.* 2022), combined with the increasing need to monitor long-term changes in biodiversity due to anthropogenic change and the decreasing costs of sequencing, has highlighted metabarcoding as an easy, scalable, and accessible method for the long-term monitoring of soil microbial diversity. However, despite the establishment of numerous large-scale, long-term soil microbial diversity monitoring initiatives (e.g., Soil Bon (Guerra *et al.* 2021)), the spatial sampling design of soil for metabarcoding-based assessments of soil microbial diversity has received comparatively little attention. For example, the European Soil Sampling Protocol provides advice based on work that predates molecular methods.

Long-term soil sampling strategies must balance scientific rigor with practical constraints, aiming to generate datasets that are comparable across temporal and spatial scales. Here, we synthesize our current efforts to find best practices for soil

sampling strategies based on empirical work. We focus on optimizing sampling intensity and spatial distribution within a plot, determining how sample intensity, placement, volume, and DNA extraction methods affect the detection of soil biodiversity through metabarcoding across a range soil biota.

Metabarcoding

Metabarcoding refers to the use of DNA marker genes to characterize the taxonomic composition of environmental samples (Taberlet *et al.* 2012). Metabarcoding has become an increasingly attractive alternative for the identification of soil microbes and microfauna, which are not easily extracted from the soil matrix (Pawlowski *et al.* 2020). Universal marker genes including the ITS region, as well as the 18S rRNA, mitochondrial 16S rRNA, and COI genes, have been used to assess the global diversity of bacteria (Thompson *et al.* 2017), fungi (Tedersoo *et al.* 2014; Větrovský *et al.* 2019), protists (Oliverio *et al.* 2020), nematodes, microarthropods (Wu *et al.* 2011), and rotifers (Robeson *et al.* 2011) at greater speeds and increasing resolution compared to isolation-based taxonomic identification. These markers allow researchers to study several groups simultaneously, filling gaps in soil biodiversity data.

The process of surveying soil biota through metabarcoding can be broadly divided in four steps 1) soil sample collection, 2) sample preparation and DNA extraction, 3) PCR and sequencing, and 4) data processing and bioinformatics (Figure 1). This report primarily focuses on the first two steps as, once DNA extracts are archived, they can be resequenced. An in-depth exploration of the third and fourth points, in particular with regards to biodiversity data interpretation, are available in (Jurburg *et al.* 2021) and (Jurburg *et al.* 2022).

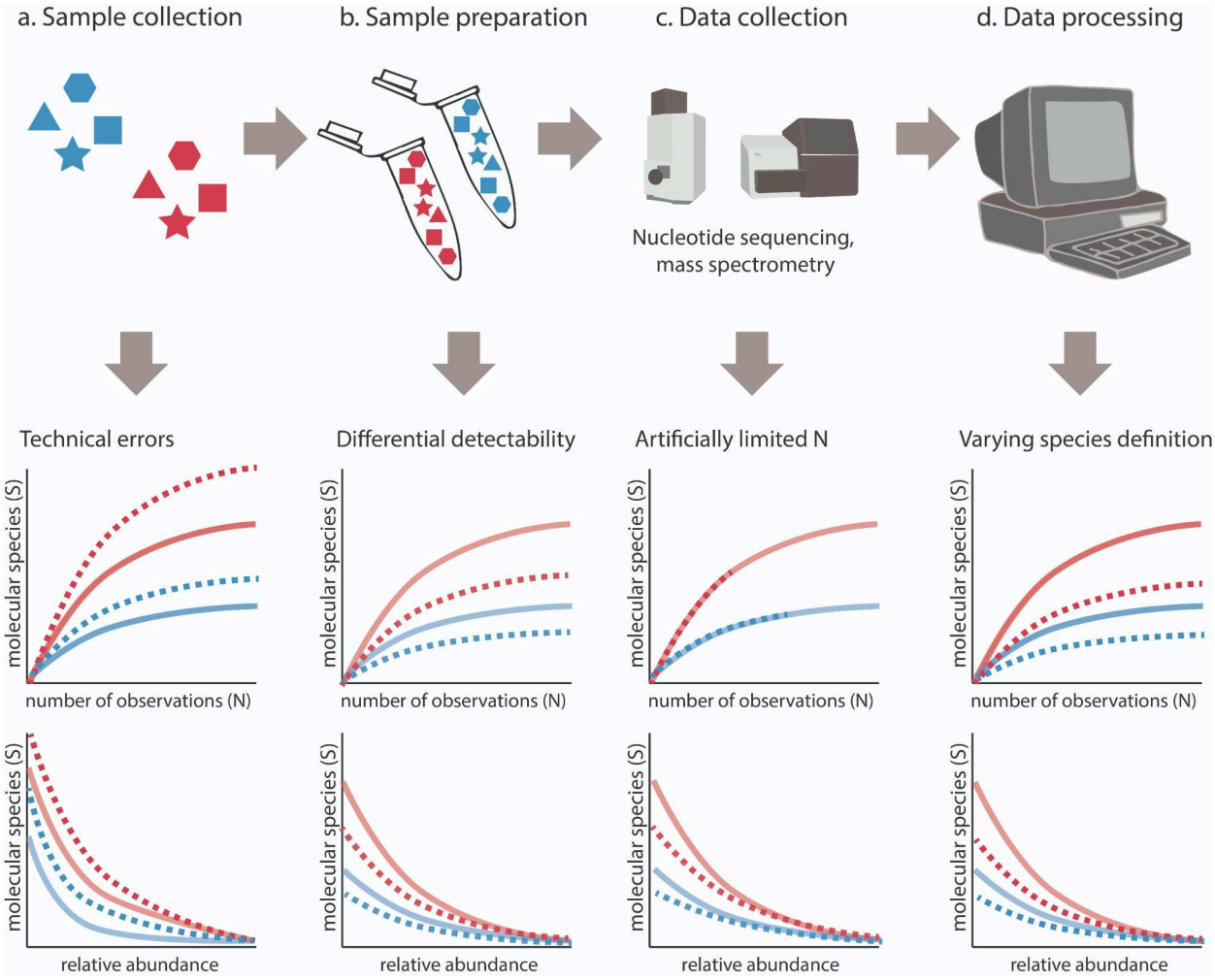

Figure 1. From Jurburg et al., 2022. Sample collection and preparation, data collection, and post-processing are inextricably linked in molecular techniques and can all potentially affect estimates of diversity. The effect of researcher choices on the number of species (S), number of observations (N) and species abundance distributions during data generation is shown below each step. The true diversity in two samples is shown in red and blue, and the measured diversity is shown as dotted lines. Technical errors during sample collection and storage can increase S (e.g., due to non-specific contamination), resulting in higher estimates for S and steeper SAD (a). In contrast, sample preparation can reduce the detectability of certain molecular entities (e.g., during PCR amplification in metabarcoding) resulting in a lower S and flatter SAD (b).

Sampling

The spatial scale of sampling (Dungan *et al.* 2002) is characterized by the volume or area of samples taken (their grain size), the spatial extent of a study, and the distance between samples (Figure 2). All three aspects of spatial scale are seldom documented in studies of soil biota, and the homogenization of multiple, randomly selected samples

within a plot is common (Jurburg *et al.* 2021). The distance decay and taxa area relationships describe ecological patterns, but have implications for how sampling designs may confound ecological patterns.

The **distance decay relationship (Nekola & White 1999)** states that as the distance between two samples increases, the similarity in composition between them decreases. This is expected for soil biota, and has been shown for terrestrial bacteria (Martiny *et al.* 2011; Goldmann *et al.* 2016). The decay of similarity with increasing distance is expected due to abiotic selection (i.e., environments become more dissimilar over space, selecting for different communities), dispersal, and drift. The DDR can become steeper when organisms are dispersal-limited, and these limits can be dictated by organismal physiology, or by environmental barriers, which are subject to anthropogenic change (Soininen *et al.* 2007; Hanson *et al.* 2012). *However, from a technical perspective, the DDR is also affected by the study extent, and steeper decay relationships are expected from smaller sample extents (Nekola & White 1999). Sampling methodologies can also have strong effects on the DDR (Clark et al. 2021).*

Related to grain size, the **taxa area relationship** (TAR) describes the increasing diversity expected with increasing sample sizes (Connor & McCoy 1979). Due to technical limitations on the study grain, sample size in soil metabarcoding research is generally determined by sampling extent. However, given the wide range of body sizes among soil biota (Jurburg *et al.* 2021), and the expectation that body size positively correlates with random variation in community structure (Zinger *et al.* 2019), it is crucial to understand how DNA extraction protocols that require different initial sample sizes inflate variance in the resulting biodiversity data.

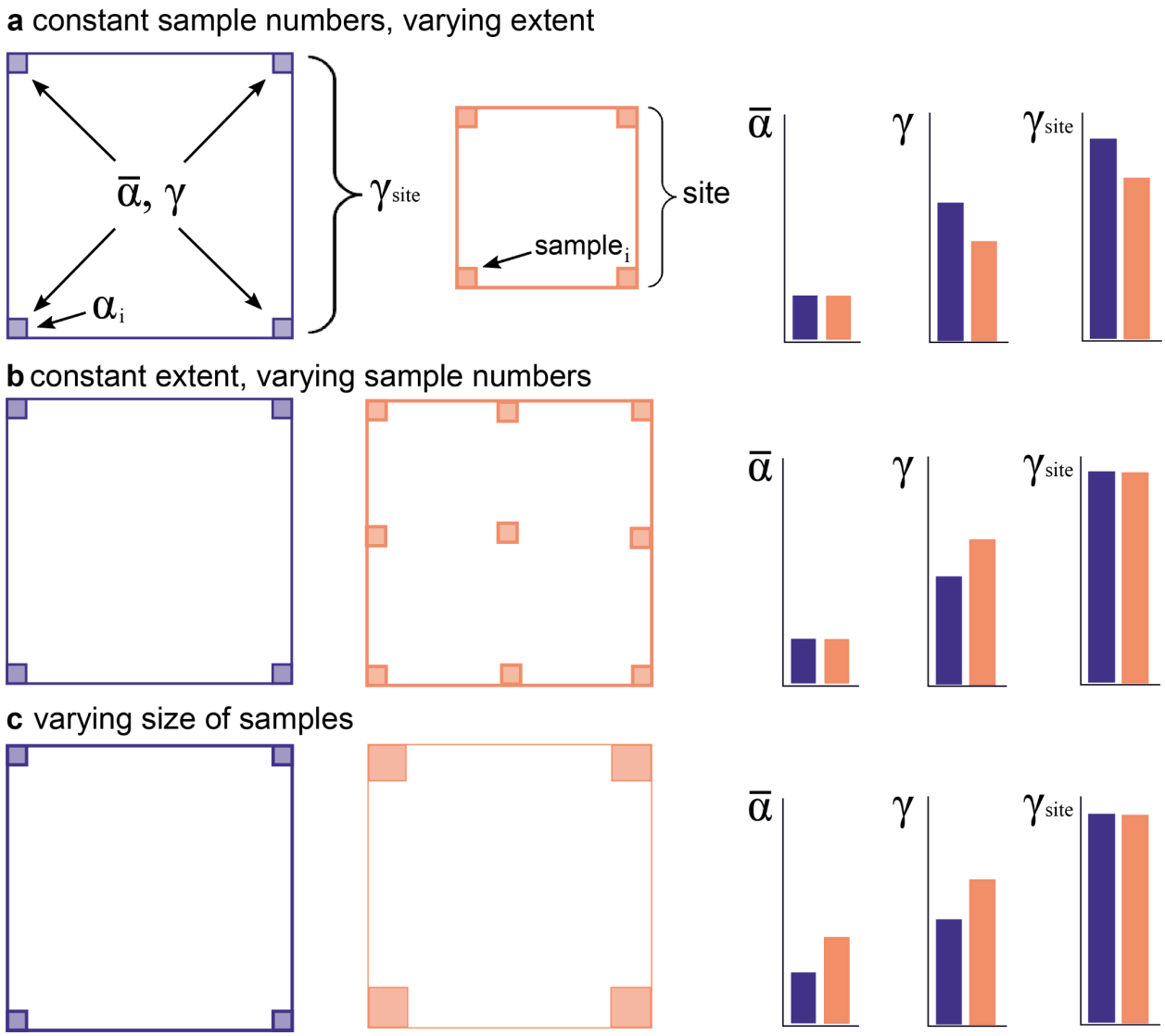

Figure 2. Adapted from Jurburg et al., 2021. Spatial sampling issues that affect average species richness per sample $\bar{\alpha}$ in a plot, total richness across all samples in a plot (γ), and total, true richness of an entire plot, i.e. both within and outside of the samples (γ_{site}). All of the expected effects stem from two ubiquitous empirical patterns: the increase of number of taxa with increasing area (taxa-area relationship) and distance decay of similarity.

Empirical work

Sample extent and sample number

With the aim of understanding the effect of sample pooling on metabarcoding-driven soil biodiversity estimates, we sampled the high intensity research grassland and forest plots (VIPs) of the three Biodiversity Exploratories sites (Hainich, Alb, and Schorfheide) as part of the SpaceMic project. These plots have been historically sampled every three years since 2011. For the soil microbiome sampling campaign, 20 x 20m and 40 x 40 m subplots are established in the grassland and forest plots, respectively. A cross-transect is sampled with a 5 cm auger, with 7 North-South and 7 West-East samples taken within the transect, from the top 10 cm of soil. This composite sample is homogenized by sieving, and a subsample of <1 g is used for DNA and RNA extraction. Together with the 2023 soil sampling campaign, the SpaceMic project took subsamples of each core (14 per plot), and performed DNA extraction and 16S rRNA and 18S rRNA gene metabarcoding on the 14 single samples and the composite sample for each plot. DNA from the samples was extracted using Macherey Nagel's NucleoSpin® Soil kit, and the 16S rRNA and 18S rRNA genes were sequenced using the Earth Microbiome Project protocols (Thompson *et al.* 2017). Reads were rarefied to 15829 reads per sample for 16S rRNA gene data and to 12000 for 18S rRNA gene data. With this design, it is possible to measure alpha diversity in each single sample, average gamma diversity for a set of single samples (equivalent to a composite sample), and the gamma diversity for the plot, and compare them to the composite sample as sampling extent (Figure 2a) and sample number (Figure 2b) are modified. Here, gamma diversity provides an empirical measure of the expected plot level diversity as the number of samples or extent per plot grows, in the absence of limitations imposed by read depth or rarefaction.

To test the effect of a growing sample extent, the cross transect design of the Biodiversity Exploratories was subsampled to create three squares of 6x6, 12x12 and 18x18 m for Grasslands and 12x12, 24x24 and 36x36 m for forest plots, labeled here as small, medium, and large extents.

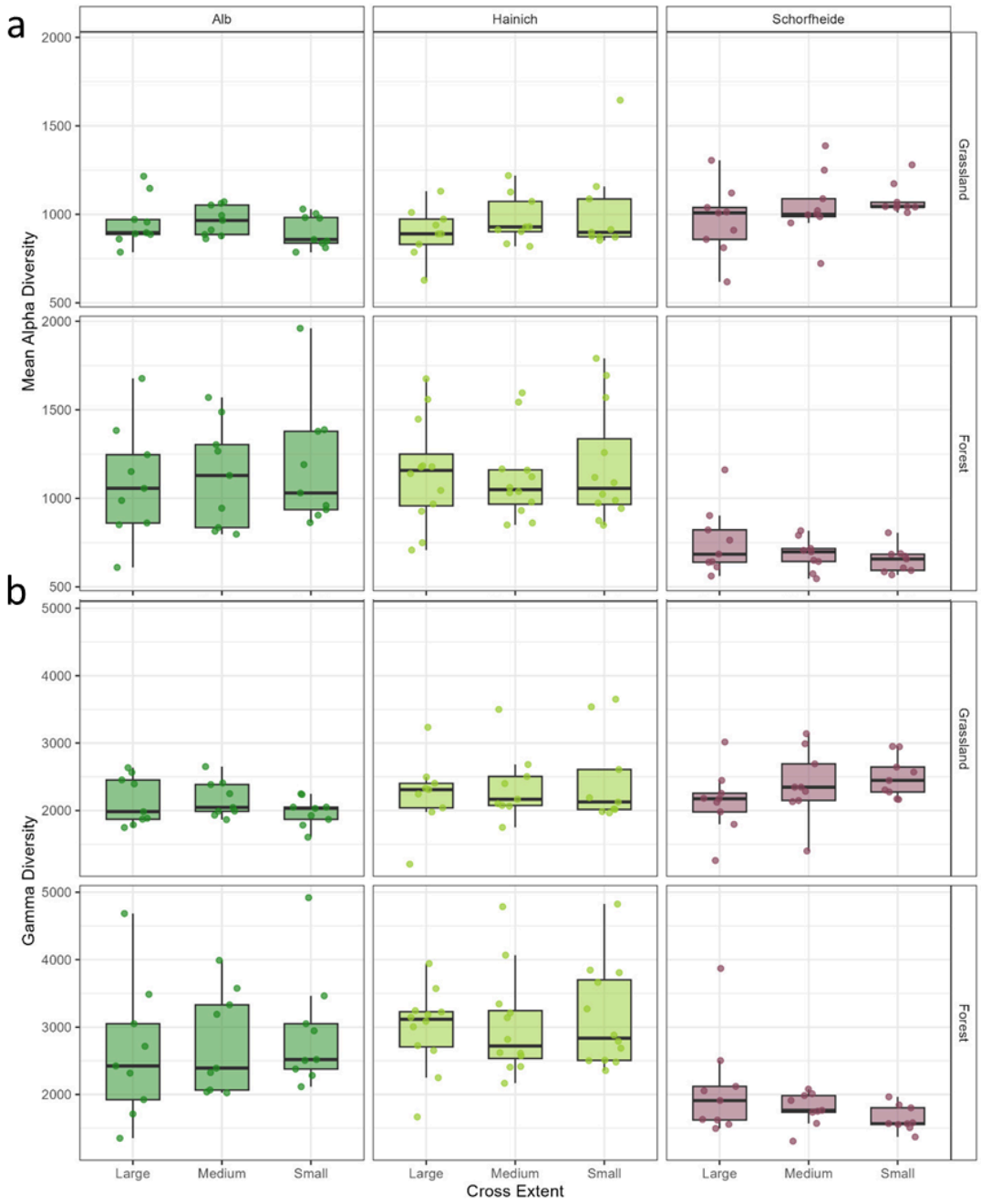

Figure 3. Alpha (a) and gamma (b) diversity in Exploratories forest and grassland plots across a range of sample extents, assessed as the bacterial ASV richness of the 16S rRNA gene. The cross transect design of the exploratories allows for the measurement of squares of 6x6, 12x12 and 18x18 m for grassland and 12x12, 24x24 and 36x36 m for forest plots, labeled here as small, medium, and large extents. Each measurement presented here considers four initial samples to define the sample extent. No differences in alpha or gamma diversity estimates were found in response to sample extent for any exploratory.

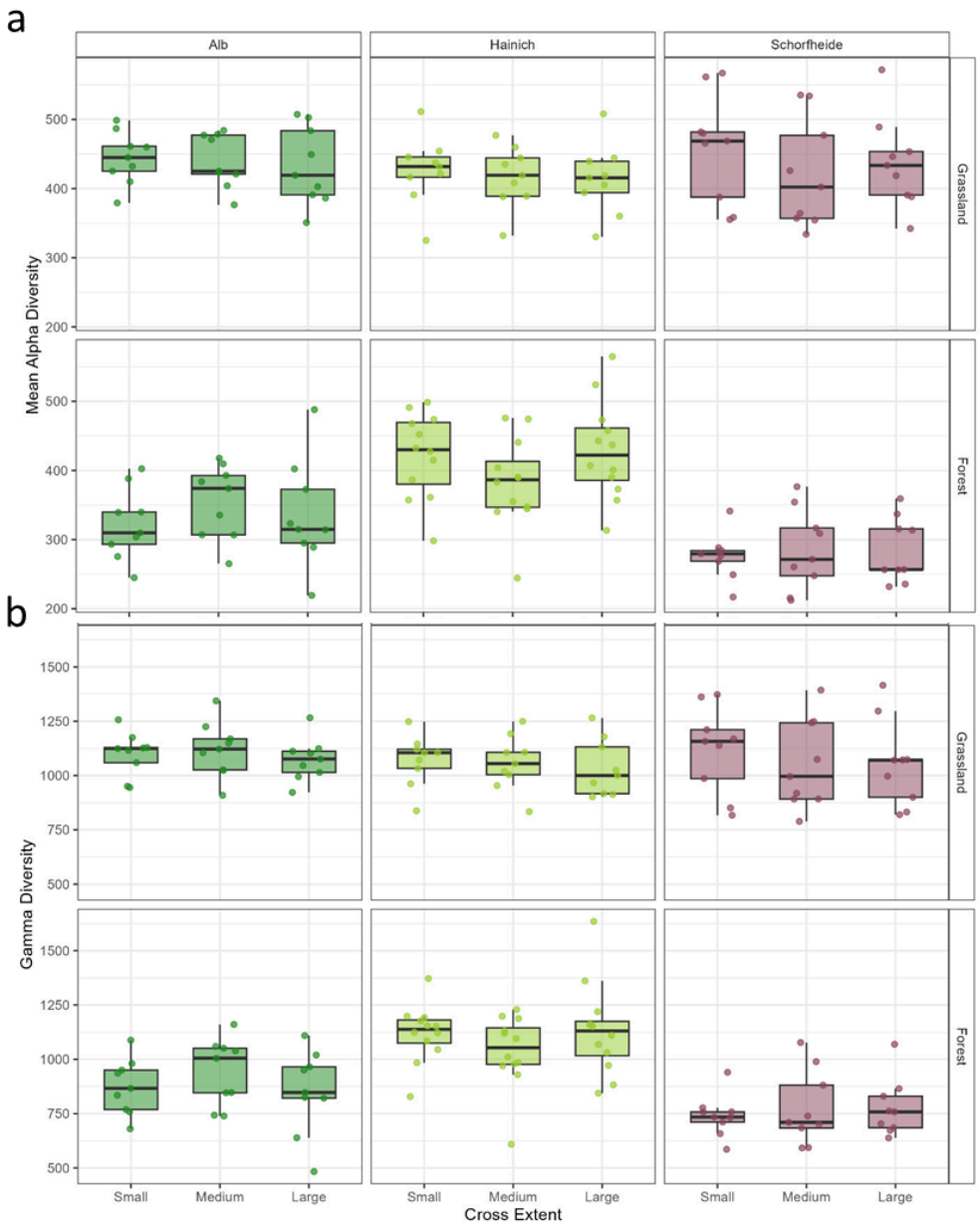

Figure 4. Alpha (a) and gamma (b) diversity in Exploratories forest and grassland plots across a range of sample extents, assessed as the eukaryotic ASV richness of the 18S rRNA gene. The cross transect design of the exploratories allows for the measurement of squares of 6x6, 12x12 and 18x18 m for grassland and 12x12, 24x24 and 36x36 m for forest plots, labeled here as small, medium, and large extents. Each measurement presented here considers four initial samples to define the sample extent. No differences in alpha or gamma diversity estimates were found in response to sample extent for any exploratory.

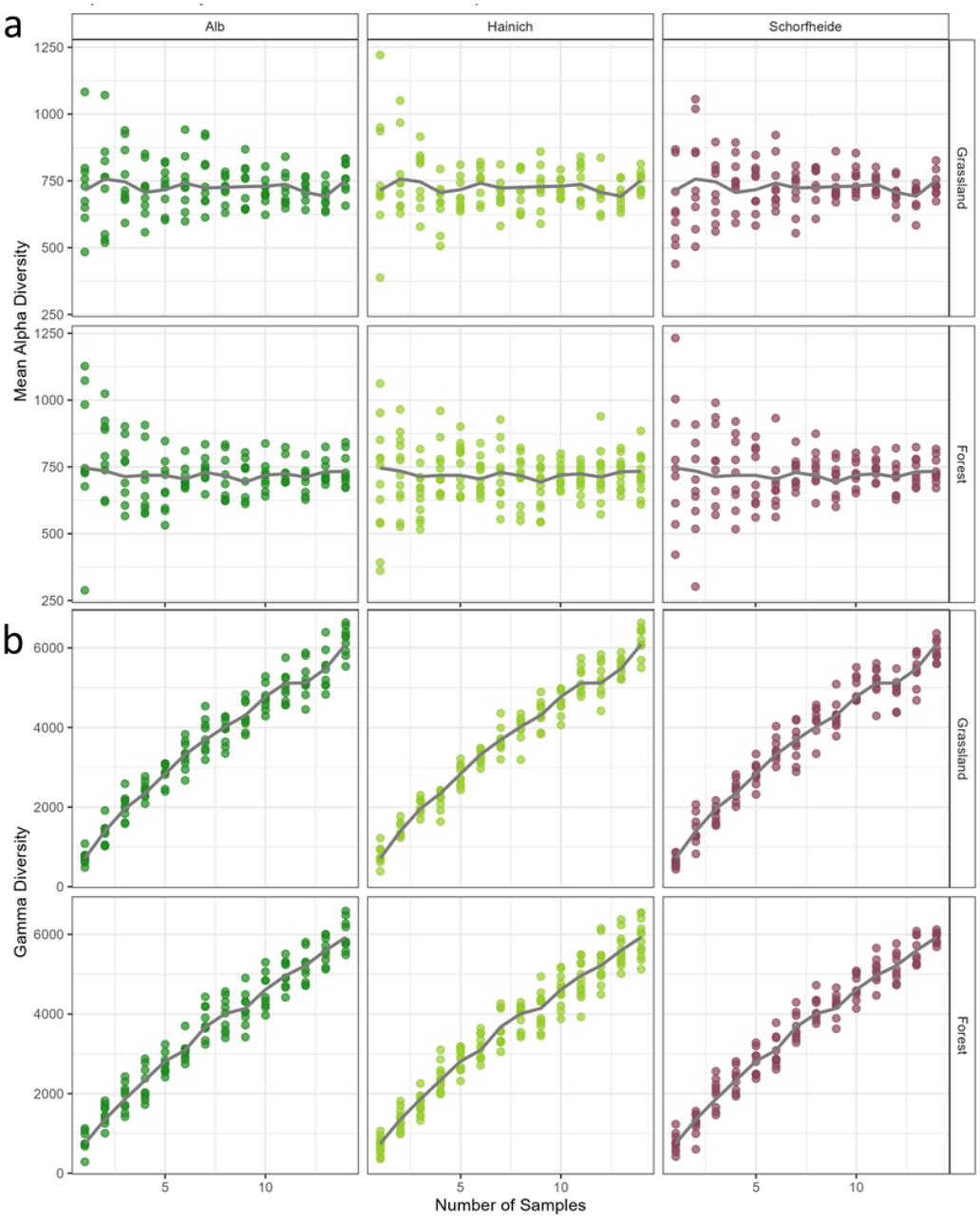

Figure 5. Alpha (a) and gamma (b) diversity in Exploratories forest and grassland plots across a range of sample sizes, assessed as bacterial ASV richness of the 16S rRNA gene. Here, each point represents a plot, and the line indicates the mean across plots for each sample number. For each sample number, individual samples were taken randomly from the plot. For alpha diversity, the average richness across these samples was calculated for each plot, and for gamma diversity, the cumulative richness for these samples was calculated for each plot. While gamma diversity was significantly, positively, and strongly correlated with the number of samples (Spearman's rank correlation, $p<0.001$ and $rho>0.97$), no relationship was found between sample number and average alpha diversity. Note that gamma diversity approximates the diversity expected at the plot level, but differs from composite samples in the depth of

observations: in a composite sample, the number of observations (i.e., reads) will stay fairly constant regardless of the number of cores that are included in the composite sample, while in our calculation of gamma diversity, the number of observations scales linearly with the number of cores included in the gamma diversity estimate.

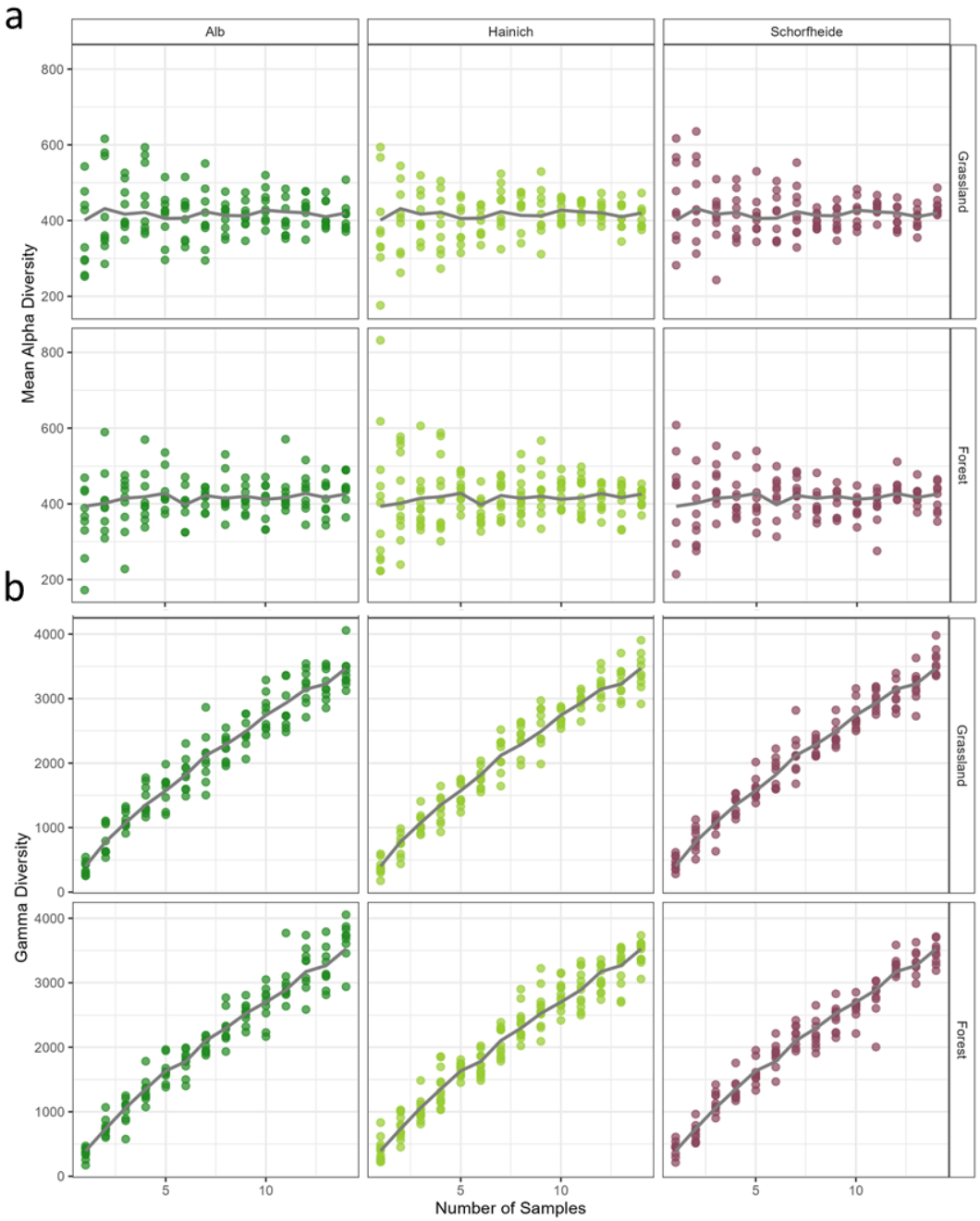

Figure 6. Alpha (a) and gamma (b) diversity in Exploratories forest and grassland plots across a range of sample numbers, assessed as eukaryotic ASV richness of the 18S rRNA gene. Here, each point represents a plot, and the line indicates the mean across plots for each sample number. For each sample number, individual samples were taken randomly from the plot. For

alpha diversity, the average richness across these samples was calculated for each plot, and for gamma diversity, the cumulative richness for these samples was calculated for each plot. While gamma diversity was significantly, positively, and strongly correlated with sample number (Spearman's rank correlation, $p<0.001$ and $rho>0.96$), weak relationships between average alpha diversity and sample numbers were found in Hainich and Alb ($0.05> p > 0.01$; $rho<0.14$). Note that gamma diversity approximates the diversity expected at the plot level, but differs from composite samples in the depth of observations: in a composite sample, the number of observations (i.e., reads) will stay fairly constant regardless of the number of cores that are included in the composite sample, while in our calculation of gamma diversity, the number of observations scales linearly with the number of cores included in the gamma diversity estimate.

Sample size

As metabarcoding techniques become more accessible, the development of molecular protocols targeting different soil organisms is flourishing, however given the body size differences among soil biota, understanding how different sample sizes capture soil biodiversity patterns (Figure 2c) is essential. Nematodes are difficult to characterize by metabarcoding due to their multicellularity, rapid mutation rate, and high copy number of marker genes (Jurburg *et al.* 2021). However, molecular methods are faster, scalable, and may offer a higher resolution of nematode biodiversity than classical isolation-based methods. Within the scope of the PhytOakMeter project, we have compared nematode diversity in an experimental plot at the UFZ Research Station Bad Lauchstaedt, Germany resulting from a wide range of DNA extraction protocols that allow for different initial soil sample masses (0.5-10g). Twelve different DNA extraction methods (a-l, Figure 7) were tested for small (0.5 g) and/or large soil samples (10 g) in quintuplicate, focusing on the effectiveness of different lysis buffers, bead beating times, and enzymatic digestion applications (Figure 7). Methods a and c were direct applications of two commercial kits, NucleoSpin® Soil kit (Macherey-Nagel, Germany) and FastDNA™ SPIN Kit for Soil, 50 mL tubes (MP Biomedicals, USA), according to the manufacturer's instructions. In method b, the reagent volumes were proportionally increased to accommodate the use of 10 g of soil, compared to the standard 0.5 g typically used. Method d was a modified protocol for extracting extracellular DNA from large soil sample volumes (Zinger *et al.* 2016). Methods e to l were customized protocols implementing mechanical bead beating and the chemical agent sodium dodecyl sulfate (SDS), with or without Proteinase K lysis, followed by DNA purification and elution using the NucleoSpin® Soil Kit. Total DNA was extracted from soil samples and sequenced with a nematode-specific 18S rRNA primer set (Nemf/18Sr2b modified to bypass the semi-nested PCR step, (Sapkota & Nicolaisen 2015)) on an Illumina MiSeq platform. Only ASVs confirmed as nematode segments (according to the PR2 database) were preserved. Then, samples were rarefied to 200 observations per sample.

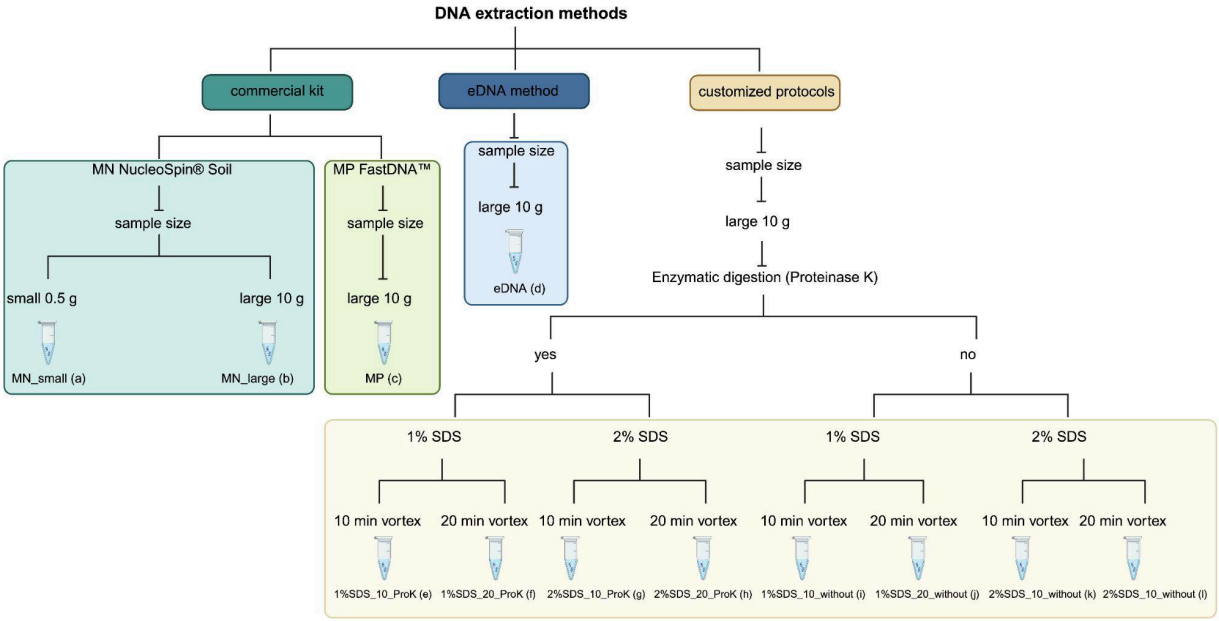

Figure 7. Experimental design. 12 different DNA extraction protocols were tested on the same soil, extracted in quintuplicate from the UFZ experimental site in Bad Lauchstaedt,Germany.

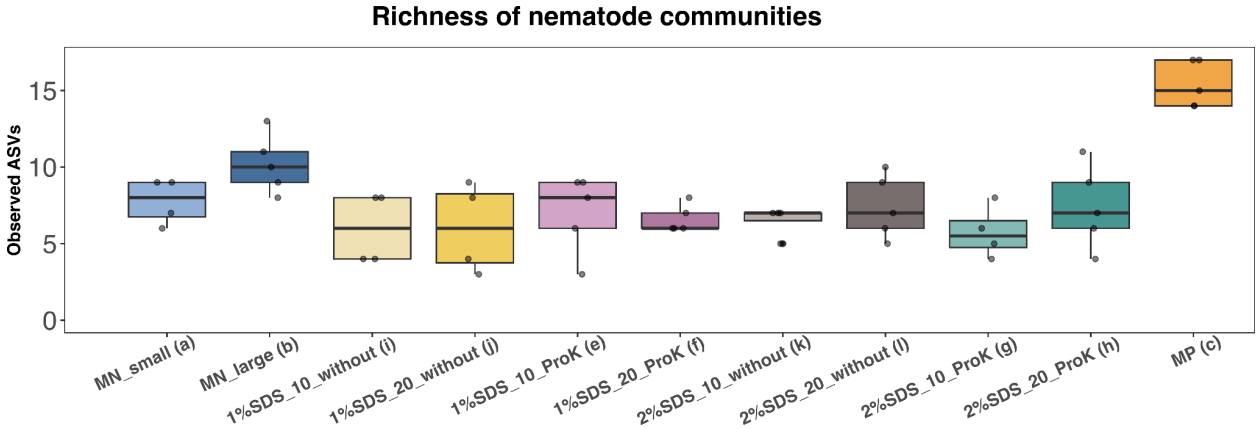

Figure 8. Observed ASV richness of soil samples processed with different DNA extraction methods. DNA extracted by MP FastDNA™ Kit (c) resulted in significantly higher nematode alpha diversity (Kruskal-Wallis, p = 0.009) compared to other methods. However, according to Dunn's test, the richness of DNA extracted with the MP FastDNA™ Kit (c) did not differ significantly from that obtained using the MN NucleoSpin® Soil Kit (a and b), regardless of the soil sample size used.

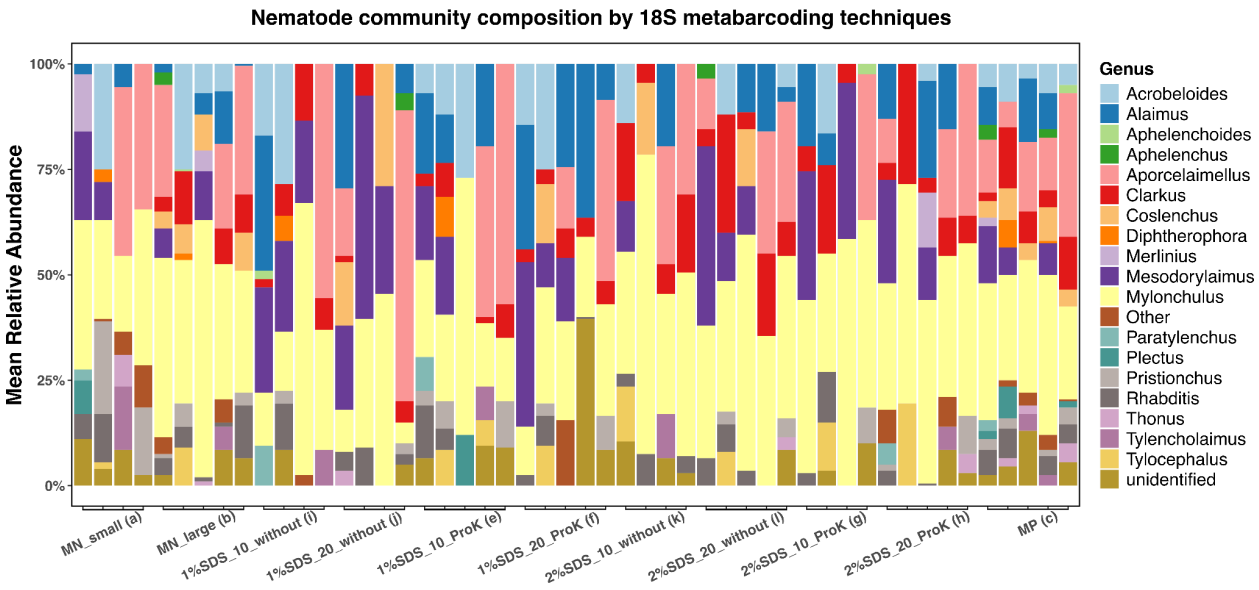

Figure 9. Relative abundance of nematode genera in soil nematode communities. Each bar represents one soil sample. *Mylonchulus*, *Aporcelaimellus*, and *Mesodorylaimus* are the most prominent nematode genera according to 18S rRNA gene metabarcoding results. A test for homogeneity of dispersions of Bray-Curtis dissimilarities found no effect of the DNA extraction method on the variance of community composition data (p = 0.176).

Recommendations

To ensure that metabarcoding surveys capture the complexity of belowground biodiversity, it is crucial to implement carefully designed soil sampling protocols that address spatial structure.

Our findings highlight that the number of soil samples from a plot included in a composite sample should result in a linear increase in the composite sample's richness, for both prokaryotes and eukaryotes. However, as the number of observations derived from metabarcoding data do not increase linearly, we expect composite samples to be less completely characterized by the method. In other words, given a limited sequencing depth, sample completeness should decrease with the number of samples included per plot. Interestingly, given a composite sample of four samples per plot, we find no effect of sample extent across the extents tested. Finally, we find that for soil fauna, the inclusion of larger sample sizes results in a slightly higher detected species richness; however the composition of the dominant portion of the community remains unaffected.

We advocate for a nested design with 1) a composite sample obtained from systematic sampling across a transect, and 2) a single sample obtained from the center of the transect. These two samples (composite and single) can undergo DNA extraction and sequencing separately, with multiple benefits. First, the composite sample follows

existing sampling designs, and allows for the inclusion of a larger amount of real biological variability from the plot, while the single sample has a definite sampling grain and extent. We highlight that due to the high heterogeneity and diversity in soil and limitations on sequencing depth, the composite sample sacrifices sample completeness (Chao *et al.* 2020) in favor of representing the diversity of the whole plot, while the single sample sacrifices representativity in favor of sample completeness and a defined sampling extent and grain.

The inclusion of two sampling scales from the same plot allows deeper ecological questions to be addressed through the inclusion of a standardized pseudo-gamma diversity measure in the composite sample relative to the single sample (e.g., homogenization). Furthermore, if the location of the sampling transect is ever compromised, this design allows for the continued sampling of the single samples, without compromising comparability. The two sample types can be used for different purposes: composite samples for broad-scale surveys, and single samples for fine-scale questions.

Relative to the established Soil Bon soil collection protocols (Figure 10), our findings indicate that sampling extent should be selected to include all eLTER plots, and that the higher the number of samples mixed into the composite sample, the more incomplete the detection of biodiversity will be for that sample. Our work also highlights the importance of maintaining the same number of samples within a composite across all eLTER sampling campaigns. The Soil Bon sampling design is therefore suitable for the composite sampling, however smaller extents and sample numbers may facilitate consistent sampling in the long-term, and may be more robust to changes.

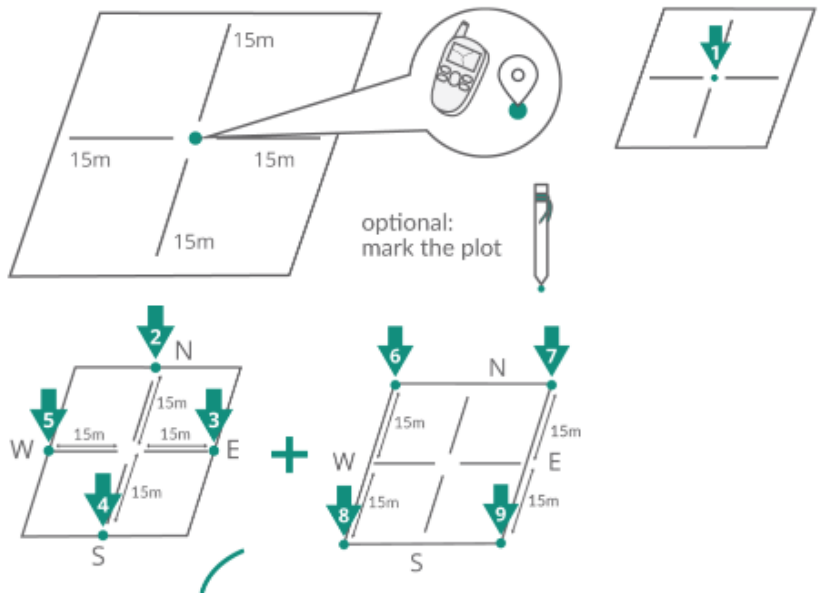

Figure 10. Soil Bon soil collection protocol, obtained from the Soil Bon Field Guide. This sampling design recommends taking a composite sample of 9 soil cores from a plot, taken at the four corners of the 30 x 30 m plot, the four middle points of the sides, and one central core. This design requires the plot to be firmly nested within a larger area that is similar to the plot, otherwise the sample will reflect border conditions.

Furthermore, any reduction of the area affects sample extent, and likely also the diversity detected, affecting comparability of samples

Based on a preliminary analysis, the Macherey Nagel DNA extraction kit can be used according to standard protocols for the metabarcoding of different groups of soil biota without significant increase in variability or decrease in the detection of diversity relative to larger sample sizes. As all sequencing is destructive, using the same DNA sample to target different groups of organisms for metabarcoding allows for the confirmation that all organisms in a sample co-occurred (this does not confirm interactions, however). Using the same DNA for different targets also reduces biases associated with DNA extraction, especially when several groups of organisms are analyzed within the same framework. Furthermore, DNA extraction and quality control is relatively labor intensive, and using the same sample for different metabarcoding targets reduces overall costs. An added benefit of this kit is that it is produced in Europe, reducing shipping costs and shipping delays.

Soil microbial metabarcoding offers an extraordinary window into the hidden diversity of belowground life. Yet, extracting meaningful insights from molecular data requires careful attention to sampling design, particularly in relation to spatial heterogeneity. By aligning methods with established ecological concepts, careful sampling designs can increase the rigor, reproducibility, and interpretability of the findings.

Bibliography

Averill, C., Anthony, M.A., Baldrian, P., Finkbeiner, F., van den Hoogen, J., Kiers, T., *et al.* (2022). Defending Earth's terrestrial microbiome. *Nat. Microbiol.*, 7, 1717–1725.
Chao, A., Kubota, Y., Zelený, D., Chiu, C., Li, C., Kusumoto, B., *et al.* (2020). Quantifying sample completeness and comparing diversities among assemblages. *Ecol. Res.*, 35, 292–314.
Clark, D.R., Underwood, G.J.C., McGenity, T.J. & Dumbrell, A.J. (2021). What drives study‐dependent differences in distance–decay relationships of microbial communities? *Global Ecology and Biogeography*, 30, 811–825.
Connor, E.F. & McCoy, E.D. (1979). The Statistics and Biology of the Species-Area Relationship. *Am. Nat.*, 113, 791–833.
Dungan, J.L., Perry, J.N., Dale, M.R.T., Legendre, P., Citron-Pousty, S., Fortin, M.J., *et al.* (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, 25, 626–640.
Fierer, N. & Jackson, R.B. (2006). The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci USA*, 103, 626–631.
Gantner, S., Schmid, M., Dürr, C., Schuhegger, R., Steidle, A., Hutzler, P., *et al.* (2006).

In situ quantitation of the spatial scale of calling distances and population density-independent N-acylhomoserine lactone-mediated communication by rhizobacteria colonized on plant roots. *FEMS Microbiol. Ecol.*, 56, 188–194.
Goldmann, K., Schröter, K., Pena, R., Schöning, I., Schrumpf, M., Buscot, F., *et al.* (2016). Divergent habitat filtering of root and soil fungal communities in temperate beech forests. *Sci. Rep.*, 6, 31439.
Guerra, C.A., Bardgett, R.D., Caon, L., Crowther, T.W., Delgado-Baquerizo, M., Montanarella, L., *et al.* (2021). Tracking, targeting, and conserving soil biodiversity. *Science*, 371, 239–241.
Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C. & Martiny, J.B.H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.*, 10, 497–506.
Jurburg, S.D., Buscot, F., Chatzinotas, A., Chaudhari, N.M., Clark, A.T., Garbowski, M., *et al.* (2022). The community ecology perspective of omics data. *Microbiome*, 10, 225.
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.
Martiny, J.B.H., Eisen, J.A., Penn, K., Allison, S.D. & Horner-Devine, M.C. (2011). Drivers of bacterial beta-diversity depend on spatial scale. *Proc Natl Acad Sci USA*, 108, 7850–7854.
Nekola, J.C. & White, P.S. (1999). The distance decay of similarity in biogeography and ecology. *J. Biogeography*, 26, 867–878.
Nunan, N. (2017). The microbial habitat in soil: Scale, heterogeneity and functional consequences. *Z. Pflanzenernähr. Bodenk.*, 180, 425–429.
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. *Sci. Adv.*, 6, eaax8787.
Pawlowski, J., Apothéloz-Perret-Gentil, L. & Altermatt, F. (2020). Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. *Mol. Ecol.*, 29, 4258–4264.
Prosser, J.I., Bohannan, B.J.M., Curtis, T.P., Ellis, R.J., Firestone, M.K., Freckleton, R.P., *et al.* (2007). The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.*, 5, 384–392.
Ranjard, L., Dequiedt, S., Chemidlin Prévost-Bouré, N., Thioulouse, J., Saby, N.P.A., Lelievre, M., *et al.* (2013). Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nat. Commun.*, 4, 1434.
Robeson, M.S., King, A.J., Freeman, K.R., Birky, C.W., Martin, A.P. & Schmidt, S.K. (2011). Soil rotifer communities are extremely diverse globally but spatially autocorrelated locally. *Proc Natl Acad Sci USA*, 108, 4406–4410.
Sapkota, R. & Nicolaisen, M. (2015). High-throughput sequencing of nematode

communities from total soil DNA extractions. *BMC Ecol.*, 15, 3.
Soininen, J., McDonald, R. & Hillebrand, H. (2007). The distance decay of similarity in ecological communities. *Ecography*, 30, 3–12.
Taberlet, P., Prud'Homme, S.M., Campione, E., Roy, J., Miquel, C., Shehzad, W., *et al.* (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Mol. Ecol.*, 21, 1816–1820.
Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., *et al.* (2014). Fungal biogeography - Global diversity and geography of soil fungi. *Science*, 346, 1256688.
Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., *et al.* (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551, 457–463.
Větrovský, T., Kohout, P., Kopecký, M., Machac, A., Man, M., Bahnmann, B.D., *et al.* (2019). A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nat. Commun.*, 10, 5142.
Wu, T., Ayres, E., Bardgett, R.D., Wall, D.H. & Garey, J.R. (2011). Molecular study of worldwide distribution and diversity of soil animals. *Proc Natl Acad Sci USA*, 108, 17720–17725.
Zinger, L., Chave, J., Coissac, E., Iribar, A., Louisanna, E., Manzi, S., *et al.* (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.
Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., *et al.* (2019). Body size determines soil community assembly in a tropical forest. *Mol. Ecol.*, 28, 528–543.