Long-read sequencing of soil eukaryote eDNA indicates vegetation as driver of soil community structure

Embla Stokke

MSc Thesis

Department of Biosciences

University of Oslo

June 2023

Acknowledgements

First and foremost, I would like to extend a very, very big thank you to my supervisor, Ella Thoen, and co-supervisors, Inger Skrede, Håvard Kauserud and Anders Krabberød, for all your guidance, help and encouragement throughout the process. Inger and Håvard, thank you for setting me up with an amazing project even when I turned up with such a limited timeline, and for offering advice and discussion along the way. Anders, thank you for answering my endless questions about bioinformatics and for helping every time I got stuck. And Ella, thank you so much for taking me in, sharing your project and making the intimidating aspects of a project like this feel a lot less scary.

I would also like to thank everyone at OMG for always being available for advice and for creating a wonderful working environment. And thanks to Marit Bjorbækmo for helping with the early and vital days of lab work.

I would also like to extend thanks to Prof. Bente Edvardsen at UiO's Section for Aquatic Biology and Toxicology (AQUA) for donating *Prymnesium parvum* culture for the mock community samples. In addition, I would like to thank Berit Kaasa at AQUA for providing the chemical analyses of the samples, and the team at the Norwegian Sequencing Centre for their expeditious sequencing service. All participants in the Archaeorhizomycets project are also due thanks for their efforts in sample collection and for making their resources available for this study, as well as Artsdatabanken for providing funding resources.

Table of contents

Abstract

Mapping diversity patterns across environmental gradients may predict how ecosystems will respond to climate change in the future. The soil biota, including protists, fungi and micro-invertebrates, are essential components of terrestrial ecosystems but it is so far not clear how these communities will respond to climate change. By analyzing soil and litter samples collected from various climates throughout Norway, the study aimed to assess how important climate is for structuring soil microeukaryote communities compared to other factors, including vegetation and local soil properties. To encompass a wide span of taxa and increase the phylogenetic resolution, long-read amplicon sequencing (PacBio HiFi) was applied with eukaryote-specific primers to obtain continuous reads spanning a ~4500 bp region of the rDNA operon (including parts of SSU and LSU genes, and the full ITS region). Sequencing of environmental DNA from soil and litter collected at 90 sites recovered over 13 000 OTUs, most of which were Fungi, Cercozoa, Metazoa, Alveolata and Streptophyta. Beta diversity analyses show significant differences in soil and litter community structures, in addition to strong community groupings by dominant vegetation type at the sampling site. Further, community structure groups indirectly along a highland/lowland-gradient. Alpha diversity is found to be differentially responsive to environmental variables among substrates and without a clear distinction between climatic and local factors. A potential diversity threshold is observed in the sequencing data and the application of long-read sequencing as a metabarcoding mapping tool is discussed.

KEY WORDS *long-read sequencing; metabarcoding; soil microbiome; eukaryotes; diversity drivers; community; bias*

---

1. Introduction

Organic soil masses make up an essential component of most terrestrial ecosystems and contain multitudes of organisms. These organisms come in a diversity of forms and fulfil a wide range of functions, both directly in the soil and in the overarching biome. Soil communities interact with their respective ecosystems on multiple levels, including cycling of elements such as carbon and nitrogen (Zhang et al., 2019; see protist review by Geisen et al., 2018; Jiao et al., 2018; Philippot et al., 2013). These interactions also extend into the shaping of above-ground species distribution patterns (Barberán et al., 2015; Prober et al., 2015; van der Heijden, Bardgett & van Straalen, 2008) and contributing to ecosystem multifunctionality (Delgado-

Baquerizo et al., 2016; Aislabie & Deslippe, 2013). Included in the wide scope of soil microbiota we find bacteria, archaea and viruses (Thompson et al., 2017; Jansson, 2022), but also a considerable variety of eukaryotes, which include fungi, protists and metazoans (Aslani et al., 2021) – and potentially immense volumes of unmapped diversity (Locey & Lennon, 2016).

Presently, diversity research remains less extensive in microeukaryotes in comparison to prokaryotes (Keeling & del Campo, 2017), followed in turn by a bias towards complex meio-/macroeukaryote organisms (del Campo et al., 2014). However, estimates on biomass distribution on Earth indicate that the biomass, and inherently, abundance, of protists and fungi is several times greater than that of all metazoans combined (Bar-On, Phillips & Milo, 2018). In addition, when microeukaryote diversity is studied it is often in marine or freshwater communities (Burki, Sandin & Jamy, 2021), leaving soil microeukaryote diversity relatively understudied and indicated to host large proportions of uncatalogued taxa (Venter et al., 2017). Considering the prevalence and ecological relevance of soil-dwelling organisms, among them microeukaryotes, there is a persisting need for robust approaches to diversity mapping. Beyond descriptive mapping, there is also a call for identifying diversity drivers and their roles in shaping community structure. Understanding these diversity patterns in soil microbiomes may be highly useful not only in unravelling larger ecosystem dynamics, but also in practical applications such as informing sustainable soil management for agriculture. This both to preserve overall ecosystem health (Prasad et al., 2021) and with regards to ensuring food security (Aguilar-Marcelino et al., 2020; Sessich & Mitter, 2015). Also, and perhaps most critically, this understanding is key in assessing soil community responses to anthropogenic climate change and pollutants (Cavicchioli et al., 2019; Crowther et al., 2019; Dubey et al., 2019). In turn, these community responses feed back into ecosystem functions and may reshape larger scale cycles (see review by Jansson & Hofmockel, 2019). As such, there are indications that including soil microbiota in large-scale carbon and geochemical predictive models may increase model accuracy (Wieder, Bonan & Allison, 2013), further incentivizing mapping of the prevalence and distribution of soil microorganisms.

After having taken a back seat to macroscopic organisms for some time, there has been an increase in studies on microbiome diversity and ecology in recent decades (see reviews by Geisen et al., 2018; and Fierer, 2017). This may be partly due to emerging alternatives to traditional diversity mapping, as size, morphology and ecology make many of these organisms unsuitable for direct observation approaches or culturing. One method of quantifying taxa

richness and composition in soil communities is metabarcoding from environmental samples (Burki, Sandin & Jamy, 2021; Semenov, 2021; Procopio et al., 2019; Taberlet et al., 2012, Ficetola et al., 2008). This approach frequently reveals novel and sometimes unexpected diversity, especially in groups where traditional diversity mapping is challenging (Tragin & Vaulot, 2019; López-Escardó et al., 2018; Arroyo et al., 2018). Metabarcoding involves DNA isolation and amplification of a short gene region with high variability within the group(s) of interest. The resulting sequence data from this region is then clustered into operational taxonomic units (OTUs) which can give an overview of diversity and may further be cross-referenced with sequence databases for taxonomic annotation (see Taberlet et al., 2012; Creer et al., 2010). Metabarcoding provides efficient estimates of diversity as well as illumination of new or cryptic taxa presence and prevalence within samples (Venter et al., 2017). Commonly used regions for this type of sequencing include the ribosomal internal transcribed spacer region (ITS; and/or subregions therein) for fungi (Schoch et al., 2012) and the small subunit region (SSU; 18S) for other eukaryotes (Francioli et al., 2021). The caveat in using short DNA regions across wide taxonomic groups lies in that evolutionary rates and variability of certain sites may differ between taxa, possibly confounding outcomes if a single short read is the only point of reference (Tedersoo et al., 2021; Tedersoo et al., 2015; Valentini, Pompanon & Taberlet, 2009; Nilsson et al., 2008). Furthermore, issues may arise when cross-referencing OTUs with sequence databases, as all taxa may not be equally well-represented (Khomich et al., 2018; Frenken et al., 2017; Nilsson et al., 2008). These factors may put limitations on the phylogenetic informativeness of and taxonomic resolution obtained from short sequence data – particularly for organisms where there is little reference data, such as soil microeukaryotes.

As an alternative to short-read metabarcoding, third-generation long-read sequencing has emerged in recent years (spearheaded by PacBio and Oxford Nanopore Technologies). While initial approaches produced high error rates, improvements in process and correction pipelines now facilitate highly accurate sequencing of thousands of continuous base pairs, allowing circumvention of traditional short-read sequence limitations without significantly compromising reliability (Dohm et al., 2020). Doing longer reads of continuous whole or multiple gene regions is already being applied successfully to phylogeny retrieval and diversity estimation (Liem et al., 2021; Ritter et al., 2020; Tedersoo et al., 2020; Jamy et al., 2019; Martijn et al., 2019; Heeger et al., 2018). Assessing the strengths of various regions in phylogenetic informativeness has indicated that more data (longer reads) can give greater depth of classification and phylogenetic resolution when applied to environmental DNA (Pearman,

Freed & Silander, 2020; Tedersoo et al., 2020; Jamy et al., 2019; Heeger et al., 2018), rendering it particularly superior (theoretically) in questions pertaining to community structure. It may also facilitate more accurate taxonomic annotation of a larger proportion of OTUs through cross-referencing across different databases covering different regions typically used for metabarcoding (Heeger et al., 2018). As such, further investigations into the application of long-read sequencing in diversity mapping and phylogenetics could be especially useful in as-of-yet enigmatic community structures.

Norway is a country of highly heterogeneous landscapes, with a multitude of elevational and latitudinal gradients as well as long stretches of coast associated with relatively warm currents and contrasting arctic waters (Ketzler, Römer & Beylich, 2021). In addition, efforts have been made to provide high-resolution mapping of nature types, resulting in detailed records of ecological and topographical gradients across the country (Halvorsen et al., 2020; Horvath et al., 2019; Bakkestuen, Erikstad & Halvorsen, 2008). This setting is ideal for studying diversity drivers and patterns, including studies on soil communities. Sampling across Norway allows for observation of changes to soil communities along several ecological gradients and vegetation types.

To map the soil eukaryote diversity in Norway, long-read sequencing is applied to eDNA from soil and litter material from 90 sample sites across the country, producing reads that span 4500 bp and cover several relevant regions for metabarcoding. The study aims to record the gamma diversity of microeukaryotes in litter and soil, and in the process add to the growing library of data on the possibilities and pitfalls associated with long-read sequencing in community metabarcoding. In addition, the study aims to investigate potential drivers of alpha diversity and community structure across two strata of Norwegian soils, focusing on whether climatic factors (precipitation, temperature, and seasonality variables) or soil geochemistry and vegetation (carbon, nitrogen, and phosphorus content; pH; dominant vegetation and plant richness) are better predictors of diversity and composition in the soil biome in the age of global change.

2. Materials and Methods

2.1 Sample collection, sample processing and sequencing

2.1.1 Sample collection

The collection, processing, and eDNA extraction of sample materials (Materials and Methods parts 2.1.1-2.1.3) was conducted as part of a separate study and concluded some time prior to the initiation of the current study, however as the project is still ongoing, methods are reiterated here in the place of external reference. Sample material was collected in June-October of 2020 and was part of a project focusing on the fungal class of Archaeorizomycetes, in which DNA was extracted from three substrates across 144 sample sites in Norway and Norwegian territory (Dr. Ella Thoen, UiO, project ongoing; see Figure 1 for sample sites selected for current study). The original set of sample sites was selected to include a diverse scope of nature types across a wide geographic range, to represent climactic gradients, topographical variation, and territorial extremes in Norway. The *Nature in Norway* nature types (NiN) (Halvorsen et al., 2020; Halvorsen et al., 2009) were referenced to select sites, and exact plot sites were chosen within an area by selecting for homogenous and easily categorised vegetation type as well as general accessibility.

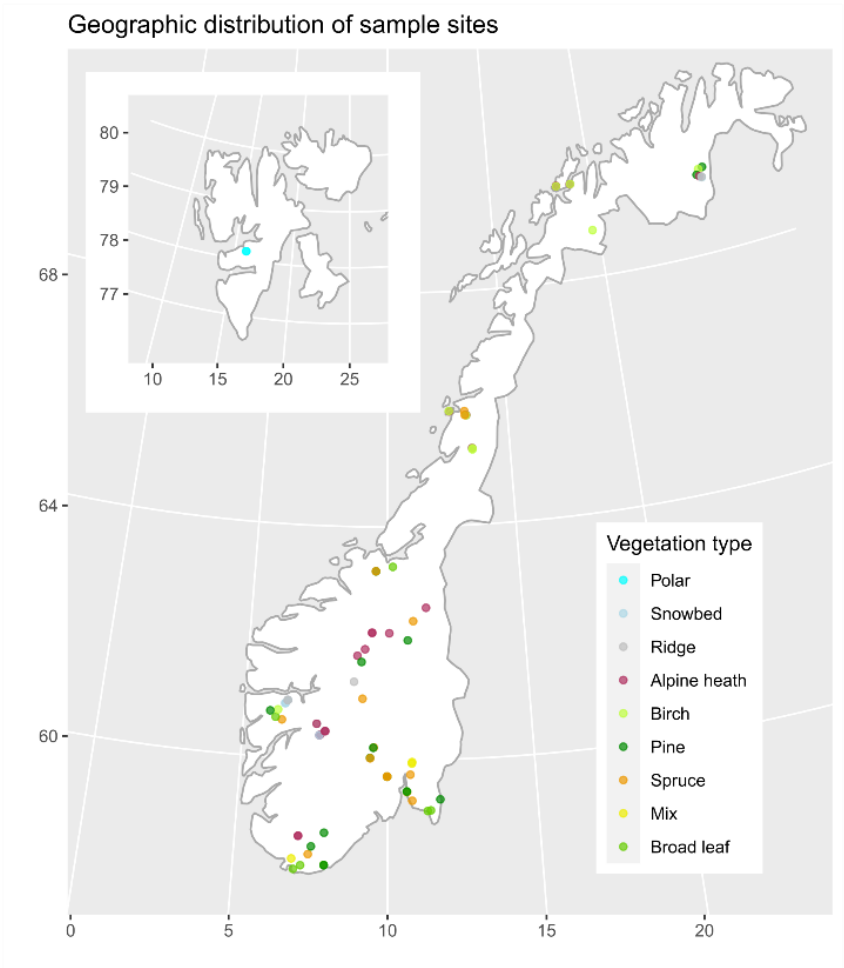

Figure 1: Map of Norway and Svalbard depicting collection sites for the study samples by their respective vegetation types.

In each sample site, a 4x4 m plot with a 1x1 m grid was measured and laid out using

measuring tape, and soil core samples were extracted centrally from each of the resulting 16 subplots. A metal cylinder measuring approx. 2.8 cm in diameter was used for coring and the depth was relative to where the mineral soil layer started or other non-organic hindrances occurred. From the core material, mineral soil was discarded and the litter separated from the organic soil. Organic soil was sifted through 5 mm sieves to separate out roots and homogenize materials and equipment was cleaned between plots using 70% ethanol and a fire torch. Materials from the 16 samples in each site were pooled in their respective substrate categories: organic soil, litter, and roots. In addition, aboveground litter was collected from the area surrounding the coring location at some sites. For the present master project, only DNA from organic soil and litter substrates were studied, and processing of root samples will not be further described. All sample material was kept cool and frozen within 24 hours to be kept at -20°C in EVOGENE lab freezers at the University of Oslo.

2.1.2 Pre-processing and DNA extraction

After thawing at room temperature, sample material was handled in Biological Safety Cabinets. For the litter material, a sterile scissor was used to cut pieces into smaller fragments of a more homogenous size and mixed by hand using sterile gloves. Similarly, organic soil material was manually mixed by hand using sterile gloves. A subsample was taken from each of the samples and contained in 50 mL Falcon tubes (tubes were filled to approx. 2/3 without a set mass) prior to freeze drying for 48 hours in a Labconco FreeZone 2.5 freeze drier (Labconco Corporation, Kansas City, MO, USA). Following freeze drying, five 6.2 mm sterile ceramic beads (M.P Biomedicals, CA, USA) were added to each sample tube for powdering and homogenizing of the material with a Fast Prep-24 beadbeater (M.P Biomedicals) by 3x40 seconds at 4.0 mills/second.

DNA extraction was performed using the E.Z.N.A. soil kit, E-Z 96 Disruptor Plate C Plus, and E-Z 96 DNA filter Plates (Omega Bio-Tek, GA, USA). 250 mg of freeze-dried, powdered sample was loaded into 2 mL Eppendorf tubes, then combined with 725 µL SLX-Mlus buffer and 725 µL CTAB. The tubes were then vortexed prior to distribution of 725 µL of the mix into the E-Z 96 Disruptor Plate C Plus. A Resch Ball Mill MM301 (Verder Scientific, Haan, Germany) at 20 rounds/second for 2.30x2.30 minutes was used for homogenizing. Following this, the E-Z 96 Disruptor Plate C Plus, E-Z 96 DNA filter, and E.Z.N.A. soil kit were applied

in combination per the manufacturer's instructions, and the resulting DNA extractions were kept at -20°C.

2.1.3 Analysis of soil chemical properties and metadata collection

Following freeze drying, each sample was subsampled to approx. 5.1 mg (median, soil) and 3.4 mg (median, litter) for measuring carbon and nitrogen content. Measurements were done in-house using a Thermo Finnigan EA 1112 Series Flash Elemental Analyzer (Thermo Scientific, Italy). For phosphorus content subsamples of 5.6 mg (median, soil) and 5.2 mg (median, litter) were measured on a SEAL AA3 HR AutoAnalyser (Seal Analytical, Germany) following the manufacturer's instructions.

In addition, subsampled material (500 mg) was suspended in distilled water for pH measurements. The diluted sample was vortexed and incubated at room temperature for approximately one hour before measuring pH using a Laquatwin-pH-11 (Horiba Scientific, Japan) following the manufacturer's instructions. For two samples, AL077 and AS109, pH measurements could not be obtained.

In addition to these measurements, geographic coordinates and elevation were also recorded, along with the NiN-type, dominant vegetation type (simplified into categories: Pine; Spruce; Birch; Broad Leaf; Ridge; Alpine Heath; Snowbed; Polar; Mix), and the number of observed plant species (plant richness). Climate data for each site was obtained as BioClimatic variables to ~1km accuracy/30 seconds from WorldClim v.2 (Fick & Hijmans, 2017, available from https://www.worldclim.org/data/worldclim21.html).

2.1.4 Sample selection and mock community assembly

From the total sample pool of 144 sites x three substrates (n = 432), 90 sample sites and two substrates (litter and soil) were selected for the current project (n = 180) (Figure 1). Roots were excluded due to the expected bias towards plant and plant-associated microorganisms. Site selection was made with the objective of preserving a full representation of nature types within forests and alpine habitats and as much geographic and ecological span as possible within a 90-site frame. Sites that appeared in geographical proximity of each other and also were of identical nature types and at similar altitudes were assessed and one (or more) were excluded.

This process was done in several rounds wherein geographical and elevational proximity became less relevant as the number of clear-cut exclusion candidates was reduced. Most NiN nature types were already represented by more than two sites and maintaining a minimum of two sites for each nature type was prioritised. No formal parameter bounds were set for exclusion, as the initial parameters had to be expanded through several rounds to reduce the number of sites sufficiently for this project.

In addition, a mock community was compiled from DNA extracted from three organisms: *Serpula similis* (Fungus), *Pleurotus ostreatus* (Fungus), and *Prymnesium parvum* (Haptophyte). Mock species were chosen based on low likelihood of occurring naturally in any of the sample material, as well as availability at the time of the project. DNA was extracted from tissue from cultures and, in the case of *P. parvum,* directly from organisms in-culture. For DNA isolation, tissues were placed in separate 2 ml Eppendorf tubes with 600 µL of 2% CTAB and 1-2 2 mm tungsten carbide beads, then ground in a Qiagen Tissuelyser II (Qiagen, Valencia, CA, USA) at 20 Hz for 2x1 minute. Following 10 minutes in a -80 °C freezer, the mixture was then incubated at 65 °C for approx. 40 minutes. After a brief cooling period, 600 µL of chloroform was added and the resulting mixture was centrifuged for 15 minutes at 13200 rpm. 400 µL from the resulting topmost layer were transferred to fresh tubes prefilled with 400 µL of chilled isopropanol and left to precipitate in a 4 °C refrigerator for 10+ minutes. Mixtures were then centrifuged for another 10 minutes at 13200 rpm before pouring out the top layer of isopropanol. 300 µL of 70-75% ethanol was then added before vortexing and centrifuging for 2 minutes at full speed. Again, the top layer of ethanol was poured out and the remaining liquid was left on an incubator plate to evaporate under the cover of sterile tissue. Sample DNA was then rehydrated with 60 µL of Milli-Q H_2O, and then the DNA concentration was measured using Qubit 4 fluorometer (Thermo Scientific, Waltham, MA, USA) following the manufacturer protocol. DNA from each sample was then combined into one suspension at concentrations corresponding to a 12:5:2 ratio for *P. parvum, P. ostreatus* and *S. similis*, respectively. All tissues/specimens for the mock community were contributed from in-house cultures (see acknowledgements). Suspended DNA and the final mock community were stored at -20 °C until further use.

2.1.5 PCR and Sequencing

Following site selection, corresponding litter and soil sample DNA suspensions were collated into new working catalogues prior to amplification (n = 180). A total of eight randomly chosen replicates from the collected samples and two mock community representative samples were also included (total sample, replicate and mock n = 190). General eukaryote primers 3nDF (forward, Cavalier-Smith et al., 2009, forward) and 21R (Schwelm et al., 2016, reverse) were used as these have previously been used successfully for long-read sequencing of eukaryotes (Jamy et al., 2019). Primers were tagged to be sample-specific, with each tag combination being used twice: once in each library, set up and contained to separate plates. For the PCR, Phusion Plus DNA Polymerase (Thermo Scientific, Waltham, MA, USA) was used in master mixes of various compositions (see Table S1 in Supplementary Materials for further details). The master mix base consisted of 4x µL Phusion Plus Buffer, 2 µL dNTP, 0.2 µL Phusion Plus Polymerase and then the remaining volumes varying between 1-3 µL BSA, 0-1 µL $MgCl_2$ and nuclease-free H_2O to reach 20 µL total. DNA input was of varying concentration but at a volume of 2 µl/sample (see Table S2 in Supplementary Materials for further details). The cycle program consisted of denaturation at 98°C for 30s, followed by 25 cycles of 98°C for 10s; annealing at 60°C for 30s; and elongation at 72°C for 5 mins. Final elongation was at 72°C for 5 mins and then products were held at 10°C until further use. PCR products were checked using agarose gel electrophoresis on 1% Agarose gels (see Table S3 in Supplementary Materials for detailed set-up). If agarose gel bands were of unsatisfactory quality, samples were run again with tweaked PCR protocols - namely by diluting sample input, increasing proportions of BSA and introducing $MgCl_2$. Most samples amplified well under similar PCR regimes, however for those that consistently amplified poorly more approaches were tested on a reduced number of samples exclusively. As a result, final products for sequencing were collected from several various PCR runs into new composite plates (Table S1 in Supplementary Materials). Bovine serum albumin (BSA) was utilised throughout the PCR variations as this has been shown to mediate some inhibitor effects associated with the soil source material (Kreader, 1996). In the latter stages of PCRs, Magnesium chloride ($MgCl_2$, Thermo Scientific, Waltham, USA) was also introduced to further facilitate amplification (as per the manufacturer's suggestion in-protocol; Koblízková, Dolezel and Macas, 1998). PCR products were cleaned and quantified at the Norwegian Centre for Sequencing (NCS) using AMPure BP beads system (Beckman Coulter, Indianapolis, IN, USA) and Qubit fluorometry (Thermo Scientific, Waltham, MA, USA) for product concentration measurement. Following this, volumes for equimolar pooling

were calculated based on product concentration measurements made at NCS, and products were pooled into two libraries which were then resubmitted to the NCS for sequencing. Each library was sequenced on one 8M SMRT cell on Sequel IIe instrument using Sequel IIe Binding kit 2.2 and Sequencing chemistry v2.0. Loading was performed by diffusion loading, with a movie time of 30 hours. CCS sequences were generated using Sequel IIe onboard analysis SMRT Link (11.0.0.146107).

2.2 Bioinformatics and statistical analyses

2.2.1 Demultiplexing, filtering, and denoising

Sequence data was returned in two separate files, which were demultiplexed and trimmed (min length 2000 bp, maximum error rate 0) separately using a linked adapter approach in *cutadapt* (Martin, 2011). Following this, the sequences were filtered (min 2500 bp, max 7500 bp, maxEE 4 and minQ 3) using the *dada2* package (Callahan et al., 2016) in R v.4.2.1 (R Core Team, 2021). *dada2* was further used for initial dereplication of sequences, denoising, and sample inference.

Chimeric detection was performed in *VSEARCH* (Rognes et al., 2016) using *uchime_denovo* with adjusted parameters (abskew 1.5, dn 0.7, minh 0.5 and xn 4.0) to minimize false negatives (at the expense of increased false positives). Only non-chimeric category sequences were exported from this step (rejecting both chimeras and borderline sequences). The *dada2* removeBimeraDenovo chimera detection tool was also tested and returned similar statistics in chimera detection (see Results), however as the borderline chimera feature is lacking it was decided to move forward with the most conservative approach of excluding both chimeras and borderlines as flagged by the adjusted *uchime_denovo*.

2.2.2 Clustering and OTU curation

Finally, the sequence data was clustered at 97% similarity using *vsearch_cluster* in *VSEARCH*. For preliminary taxonomic annotation, centroids were queried against the Protist Ribosomal Database (PR^2 v. 4.14.0; Guillou et al., 2013) using *BLAST+* v.2.13.0 (Camacho et al., 2009), retaining maximum 20 hits per query which were then sorted by bit score, e-value, and per identity match percent, retaining only the top hit for each successful query. Those hits were

then further filtered to discard Archaea hits, hits with a per-identity match of < 85%, alignment length of < 200 bp and/or match start in query sequence after the initial 10 bp. Sequences which could not be successfully matched to PR^2 database entries using this approach were excluded from further analysis, with the intention of utilizing successful hits as a proxy for definite eukaryote sequences. This was done as part of a consistently conservative approach to eliminate as much potential noise as possible from the dataset. Finally, the unclustered sequences of individual samples were mapped against cluster centroids using *usearch_global* (id 0.97) in *VSEARCH*.

A separate annotation round was made where queries returned from PR^2 as the major groupings Fungi, Streptophyta or Metazoa were rerun using BLAST+ against databases UNITE (Fungi; UNITE v. 16.10.2022; Abarenkov et al., 2022; Nilsson et al., 2018) and NCBI (Streptophyta and Metazoa; NCBI v.2.2022; Wheeler et al., 2007). Hits from all three databases were filtered to the same quality threshold as previously described, then those query sequences that remained without appropriate quality hits were run through all three databases in parallel using BLAST+. The resulting hits were combined, quality filtered again and then sorted by bit-score, e-value and per-identity match to retain only the best hit for each query prior to being joined to the first level of hits.

2.2.3 Phylogenetic inference

A reference library was assembled from SILVA SSU and LSU databases (release 138.1; Pruesse et al., 2007) by retrieving and concatenating all SSU and LSU sequences which could reliably be attributed to the same genome and were also relevant to our database annotated data. This was determined by matching accession numbers and inspecting region coordinates to indicate that results were a) of the same origin and b) from the correct region. In addition, 7 external sequences were retrieved for Amoebazoa (NCBI Assembly) and two for Archaeorhizomyces (NCBI; one continuous SSU-ITS-LSU sequence from *A. finlayi* and one sequence concatenated from the SSU and LSU sequence for *A. borealis*). Reference sequences were concatenated and aligned in Geneious (Geneious Prime version 2022.2, Biomatters; Available from https://www.geneious.com) using the MAFFT v7.490 plugin (Katoh & Standley, 2013). Sequence alignment and simple tree visualizations were reviewed and the sequence selection modified to eliminate non-eukaryotic, lacking or irrelevant reference

sequences, resulting in a total of 274 reference sequences which were subsequently joined to the centroids.

The combined full sequence library was then aligned using MAFFT v7.470 and a maximum likelihood tree assembled in FastTree v.2.1.11 (Price, Dehal & Arkin, 2010) with the general time-reversible model specified. The resulting tree and alignment were inspected for topographical errors with respect to reference sequences as well as sequence anomalies in the sample sequence data. In addition, alignment length was reduced by removing 99%, 75% and 50% gaps sequentially in Geneious to improve alignments and it was also decided to exclude the ITS regions as they are difficult to align due to frequent hypervariability. From these, new alignments and trees were produced with the previously described approach and inspected to find that the 50% gap removal along with ITS exclusion gave the best-resolved phylogeny. The final tree is composed of 246 reference sequences and 13 062 sample centroid sequences.

To annotate the phylogenetic tree, taxonomic annotations obtained through the multi-database approach (described in Materials & Methods part 2.2.2) were applied. The final tree was visualized using Interactive Tree of Life (iTOL v. 6.7.4, Letunic & Bork, 2021). Rough clades were outlined manually where possible following a visual inspection of tree topology and a consensus judgement based on the majority and placement of database-sourced taxonomic annotations and the curated reference sequences to delineate some larger groups in the eukaryote tree of life. To produce an estimate of the level of confidence that all assigned clade members were indeed coherent with a monophyletic clade, all OTUs within the presented clades were checked for proportion of "correctly" annotated OTUs ("correctly" indicating adherence to the overall clade annotation). In addition, the top 15 most overall abundant taxa (relative abundance from raw data) were calculated to phylum-equivalent rank in R v.4.2.2 using R package *fantaxtic*'s *top_taxa* function and visualized using *plot_nested_bar* from the same package. This was repeated for the top 15 most abundant taxa in litter and soil separately.

2.2.4 Statistical analyses

All statistical analyses were performed in R v.4.2.2 (R Core Team, 2021) and employed the following packages: *vegan* (Oksanen et al., 2022); *geosphere* (Hijmans, 2022); *agricolae* (de Mendiburu, 2021); *psych* (Revelle, 2023); *FSA* (Ogle, Doll, Wheeler & Dinno, 2023). Data handling was done using packages *phyloseq* (McMurdie & Holmes, 2013) and *tidyverse* (Wickham et al., 2019). Data was further curated and visualized using *ggplot2* (Wickham,

2016); *corrplot* (Wei & Simko, 2021) and *fantaxtic* (Teunisse, 2023). Map visualizations were produced with additional packages *maps* (Becker et al., 2022) and *cowplot* (Wilke, 2020). Packages *gt* (Iannone et al., 2023) and *gtsummary* (Sjoberg et al., 2021) were utilized for producing tables.

Prior to analysis, all numeric metadata variables were zero-skewness transformed and ranged (0-1) according to Økland, Økland & Rydgren (2001), except for those whose distribution could not be improved towards normality through this approach. For those variables, a simple ranging was performed using the function *decostand* from the R package *vegan* with *method = "range"* (applies to the % carbon content in the samples (C) and mean temperature of warmest quarter (°C, *bio_10*), respectively). Longitude and latitude were included both as zero-skewness transformed ranged variables and as coordinates, separately. In addition to the originally collected sample chemical properties, the carbon to nitrogen ratio in each sample was also calculated (*C:N*) and included. Metadata range summaries and transformations can be found in Table S4 in Supplementary Materials.

Out of the 19 available BIOCLIM variables, a subset of 10 was selected following correlation analyses in the form of a Pearson's product-moment correlation visualization, which indicated collinearity among several of the variables, and Principal Component Analysis (PCA) visualization. Those BIOCLIM variables that shared similar directionality and distance vectors in the PCA were assessed and a representative for a given cluster was selected, primarily prioritized by the simplicity of the variable (i.e. as close to direct observational climate data as possible). Final metadata selection can be viewed in Table 1 and correlational visualizations can be found in Supplementary Materials Figures S1 and S2.

Table 1: Environmental variables included in final analyses.

Variable	Description
SITE	Short-name for site
TYPE	Substrate; Litter or Soil
forest_type	Dominant vegetation or forest type at sample site
N	% Nitrogen i sample
C_N	Carbon-to-Nitrogen ratio, calculated from raw percentages
C	% Carbon in sample
P	% Phosphorus in sample
pH	pH measured in sample
lat	Latitude
long	Longitude
plant_richness	Plant richness per site
MAS	Metres above sea level
bio_1	Annual mean temperature
bio_2	Mean diurnal ranges (mean of monthly (max temp - min temp))
bio_3	Isothermality (BIO2/BIO7)(x100)
bio_4	Temperature Seasonality (standard deviation x 100)
bio_8	Mean Temperature of Wettest Quarter
bio_9	Mean Temperature of Driest Quarter
bio_10	Mean Temperature of Warmest Quarter
bio_11	Mean Temperature of Coldest Quarter
bio_12	Annual Precipitation
bio_15	Precipitation Seasonality (coefficient of Variation)

In the mock community samples, the three original input species were recovered successfully (according to PR^2, NCBI and UNITE annotation) – *P. ostreatus* was not returned as a good match in any of the databases, however as all but one of the high-abundance centroids could be confidently annotated as either *S. similis* or *P. parvum*, this may be due to poor coverage as the best hit for the remaining centroid was limited to class level Agaricomycetes. In addition, two OTUs annotated as Leotiomycetes were retrieved - however the collective abundances were 2 and 5 reads and so they are likely artefacts of annotation or filtering approaches. Aside from this, no unexpected OTUs were retrieved.

Sample replicates were subset and assessed for discordance. Non-metric multidimensional scaling (NMDS) plots were produced for visual inspection of sample/replicate dissimilarity distances using *metaMDS* from *vegan* and resulting ordination plots. PERMANOVAs were applied to test for dissimilarity between the original samples (as a whole) and the replicates (as a whole) (*adonis2* from *vegan*, Bray-Curtis distance and 9999 permutations) – while not speaking to the pairwise similarity, this was considered sufficient when viewed together with the NMDS plot. Sample replicates and their counterparts were examined for significant differences in community structure and were not found to be significantly dissimilar (PERMANOVA, $r^2 = 0.014$, $p = 0.96$) and so were considered satisfactorily non-indicative of major amplification or bioinformatic issues.

For the statistical analyses, mock community and replicate samples were excluded. As the resulting dataset was large and skewed but also contained a considerable amount of zero counts, it was decided to work with the data without any initial transformation and instead apply non-parametric analyses. For analyses, the dataset was processed 1) as a whole, without the exclusion of library size outliers and otherwise indicated outliers or subsetting, for initial alpha and beta diversity mapping; 2) as a whole, excluding outliers; 3) as subsets of substrate types, excluding outliers. Outliers were determined from library size (> 40 000 reads in a sample) and from initial NMDS ordinations no further outliers were indicated. When filtering to exclude size outliers (samples with > 40 000 reads), eight samples were excluded – four litter samples (AL060, AL061, AL092 and AL112) and four soil samples (AS050, AS079, AS108 and AS128). In addition to replicate and mock exclusions, two samples were lost to poor amplification (AL107 and AS116) along with one replicate (AL002-REP) which was lost due to a traceable and isolated error during PCR product pooling.

For alpha diversity indices, observed richness of OTUs, Chao1 richness estimates, Shannon-Wiener Diversity Index and Evenness scores were calculated for each sample. Chao1 Richness estimates and observed richness corresponded exactly; Shannon Diversity Index takes into account evenness; so Chao1 and Shannon Diversity Index scores were selected as alpha diversity representatives for analyses into environmental variable correlations. As Shapiro-Wilks normality tests and Kolmogorov-Smirnov normality tests almost exclusively indicated non-normal distribution of all alpha diversity indices, several transformation alternatives were applied in an attempt to approach normality (*log; log10; sqrt*). However, as normality could not be satisfactorily and consistently asserted statistically and visually (distribution histograms) by any of these approaches, non-parametric analyses were applied to alpha diversity indices as well.

To investigate if alpha diversity indices vary with categorical metadata, Kruskal-Wallis rank sum tests were applied to each categorical variable independently. Significant results ($p < 0.05$) were followed by post hoc pairwise testing using Dunn's Kruskal-Wallis multiple comparison tests with Bonferroni adjustments. Continuous variable correlations were assessed by applying Kendall's rank correlation (tau) to each variable independently. In addition, Generalized linear models (GLMs) were set up with alpha diversity indices as response variables using *glm()* and default specifications.

For beta diversity analysis, the dataset was filtered to exclude OTUs that occurred in < 2 samples and total-transformed using *vegan*'s *decostand* and *method = "total"* to yield a relative abundance table. The function *adonis2* in the package *vegan* was used to perform PERMANOVAs on Bray-Curtis distance matrix with 9999 permutations, assessing each categorical and continuous variable independently. NMDS was performed using *metaMDS* in *vegan* with Bray-Curtis distance matrix, two dimensions, and a maximum of 250 tries and visually inspected for fit using *stressplot()* and *goodness(),* also from *vegan*. The resulting ordination was used to quantify correlation vectors by applying *envfit* with 9999 permutations.

To investigate the effects of geographic distance between sample sites, a Mantel statistic test was performed using *mantel()* with 9999 permutations from *vegan* with Spearman's rank correlation and Bray-Curtis distance matrix; the geographic distance was calculated using *distm()* and function *distHaversine* from the *GeoSphere* package and non-transformed coordinates.

3. Results

3.1 Data characteristics

A total of 7 881 289 HiFi sequences across two libraries with a mean number of CCS passes of 19-20, a Q45-Q46 median quality score and an average read length of 4443-4448 bp was returned from PacBio Sequel IIe sequencing at the Norwegian Sequencing Centre. Following demultiplexing by strict tag matching, the total number of reads was reduced to 5 838 735. Filtering, denoising, chimera exclusion and data curation reduced the post-demultiplexing total read count by 34.1% and the per-sample average read retention in analysis-ready samples was 65.7%. Chimera formation rate was seemingly low in the sample data: 0.61% of unique sequences were flagged as chimeras and 12.66% as borderline category (*uchime_denovo* with adjustments). Both chimera-flagged and borderline sequences were excluded, amounting to a total of 8.13% of read abundance. The chimeric formation rate was considerably greater in the mock communities compared to rates observed in the environmental samples. Depending on chimeric detection approach (*removeBimeradenovo* in *dada2* or *uchime_denovo*) in *VSEARCH*) chimeric rates in the mock community samples averaged at 59-86% (unique sequences; 7-11.5% abundance; not including borderline categories).

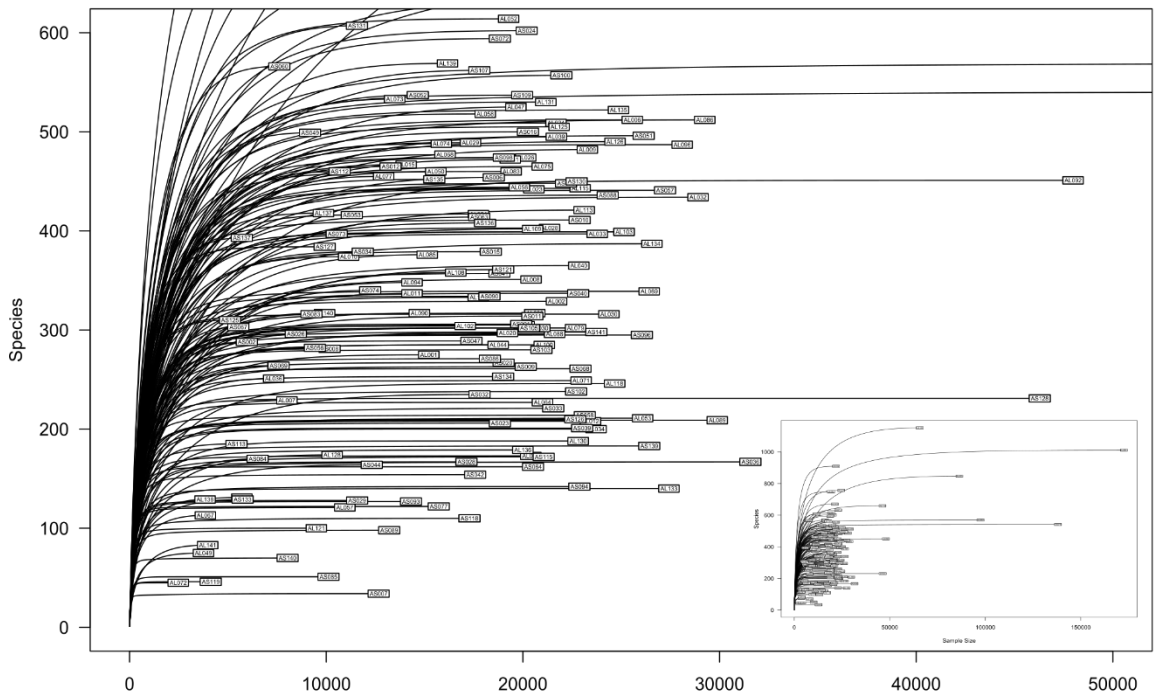

Figure 2: Species accumulation curves for samples, focused on the most common range of plateaus (100-500 OTUs across 10 000-40 000 reads). The inlay shows the curves for size outliers.

The final dataset used for statistical analysis contained 3 695 632 reads across 13 129 OTUs and 178 samples (outliers included; mock community samples and replicates excluded). Species accumulation curves were largely similar across samples and appeared to all plateau irrespective of sample size (Figure 2).

Read distribution in the overall dataset was moderately skewed, with considerable overlap between OTU commonality and total OTU abundance: the 10% most common OTUs also held 78.8% of total read abundance. Read distribution was also skewed between samples with a few samples containing considerably higher overall read counts than the majority: the 10% most abundant samples accounted for 26.6% of total reads. There was a significant correlation between sample read abundance (library size) and number of OTUs observed (Kendall's rank correlation on outlier-excluded data, $\tau = 0.158$, $p < 0.01$) and most samples fall in a middle range of 100-500 OTUs per sample (Figure 3). A considerable proportion of total OTUs occurred only once (i.e. in their respective sample and no others): 49.4% (6483) of OTUs but only 3.5% of total reads. The per-sample prevalence of unique OTUs was variable but low, with an average of 8.7% per sample.

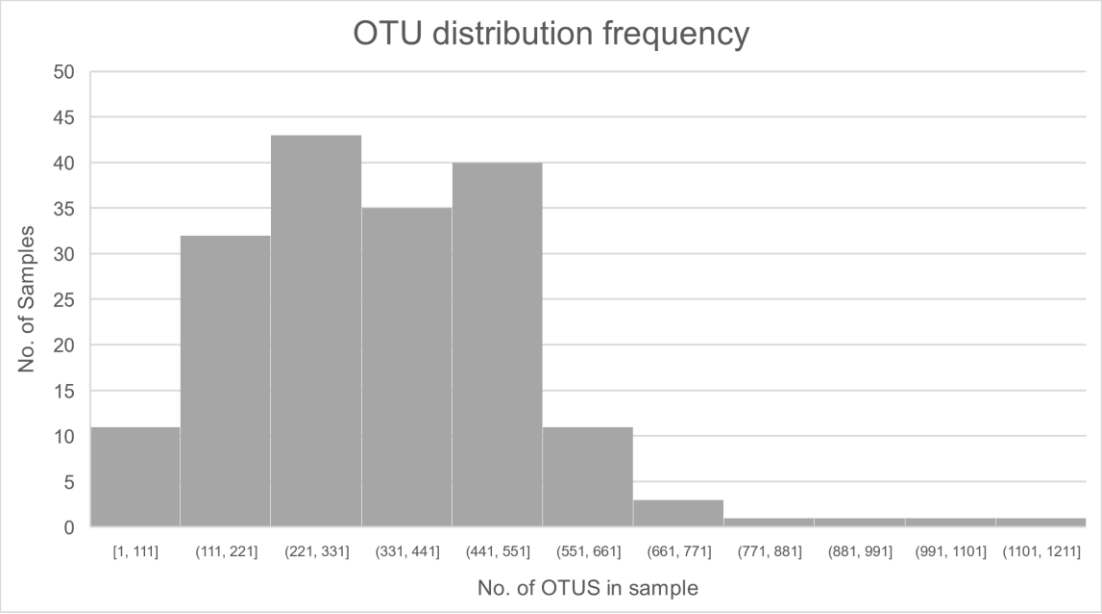

Figure 3: Histogram depicting frequency of OTU distribution across samples.

3.2 Phylogeny and taxonomic composition

Overall OTU richness is described from the phylogenetic tree, with delimited groups checked against database annotations to give an estimate of annotation accuracy relative to phylogenetic

placement. 96.87% of the OTUs, distributed across 14 groups, were checked for phylogeny-annotation coherence - the remaining OTUs were primarily scattered in smaller, paraphyletically placed groups and so were not assessed against any major retrievable groups. The assessed groups were found to have an average coherence (match percent) of 93.17% (min. 70,59%; max. 100%) between taxonomic annotation and phylogenetic placement. The five most diverse groups were: Cercozoa (~3158 OTUs; 99.97% annotation match); Ascomycota (~2682 OTUs; 98.14% annotation match); Basidiomycota (~1430 OTUs; 97.48% annotation match); Non-Dikarya Fungi (~1418 OTUs; 70.59% annotation match); and Metazoa (~1353 OTUs; 87.58% annotation match). Together, these groups made up 76.9% of the overall observed richness (Figure 4).

Abundance data is tied to the taxonomic annotation and so is, as demonstrated by the varying annotation-phylogeny cohesion, somewhat inaccurate. Working from a phylum-level rank equivalent and conglomerating relative abundance distribution in groups to match the groups mapped in the phylogeny, the 15 most abundant phyla-level groups investigated made up 96% of total abundance. This approach does leave some uncertainties as it is only based on phylum-level conglomerate abundance, however the distribution of the remaining 4% is unlikely to cause major shifts in the observed abundance scheme. Fungi dominated the relative read abundance with Ascomycota alone providing 33.8%. Then followed by Streptophyta (16%) and Basidiomycota (15.5%); Cercozoa (8.2%), Metazoa (8%), Alveolata (6.8%), and lastly non-Dikarya fungi (6.2%) and Chlorophyta (0.9%). Uncultured fungi were also in this range (0.7%), however as this annotation is not phylogenetically placeable it was included in the *other* category in some visualizations. When top taxa are extracted for soil and litter separately, distribution of relative read abundance differs between the substrates (Figure 4).

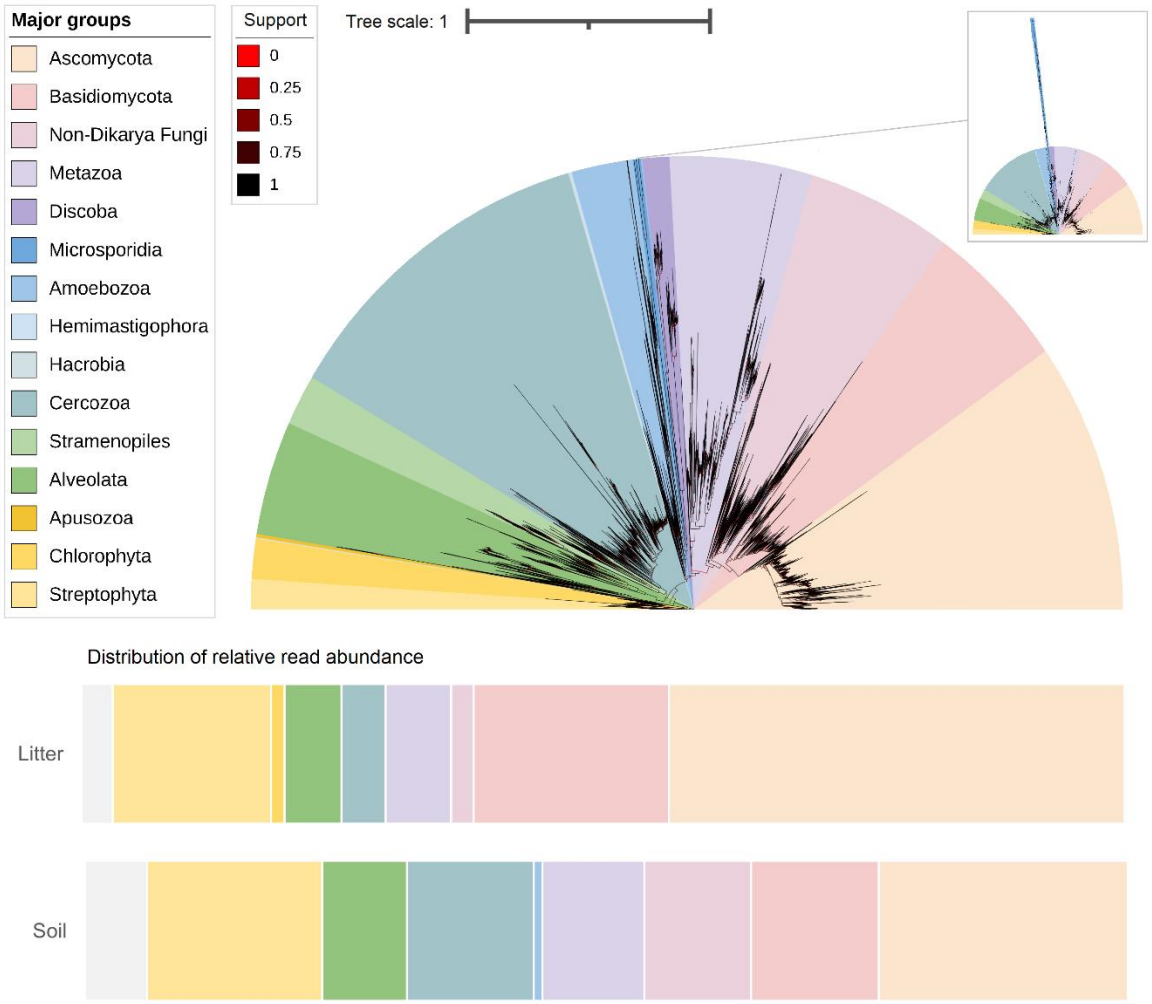

Figure 4: Phylogenetic tree from total OTUs, with major groups indicated by color and poorly supported nodes marked red. An inlay is provided for long-branch scale depiction, as well as stacked barcharts for distribution of relative read abundance in substrates separately.

3.3 Alpha diversity trends in soil eukaryotes

OTU richness and library size (sample read abundance) correlated positively (full dataset, Kendall's rank correlation, $\tau = 0.21$, $p < 0.001$) and also when outlier samples were excluded ($\tau = 0.16$, $p < 0.01$) (Figure 5). Shannon diversity index score also correlated positively with library size ($\tau = 0.16$, $p < 0.01$; outliers excluded, $\tau = 0.14$, $p < 0.01$), whereas evenness scores did not correlate significantly with or without outliers.

Figure 5: Correlation plot for Chao1 Richness Estimates and library size (sample read abundance); outliers excluded. Note that the regression line is illustrative and only representative in direction, not slope.

Percent unique OTUs correlated positively with Shannon Diversity index scores ($\tau = 0.36$, $p < 0.001$; outliers excluded, $\tau = 0.24$, $p < 0.001$) and with Chao1 richness estimates ($\tau = 0.26$, $p < 0.001$; outliers excluded, $\tau = 0.35$, $p < 0.001$). Evenness score did not correlate significantly with percent unique OTUs, with or without outliers. Library size and percent unique OTUs in library correlated positively only when outliers are included ($\tau = 0.13$, $p < 0.05$).

Alpha diversity indices correlated moderately among themselves: Shannon Diversity index score and Evenness score correlated positively ($\tau = 0.40$, $p < 0.001$; outliers excluded, $\tau = 0.42$, $p < 0.001$), as do Shannon Diversity scores and Chao1 Richness estimates ($\tau = 0.51$, $p < 0.001$, outliers excluded, $\tau = 0.52$, $p < 0.001$). Evenness and richness estimates did not correlate significantly with or without outliers.

For alpha diversity indices and environmental variables, analyses were made on data where the previously described outliers were excluded. Tables containing further Kendall's correlation and GLM output details can be found in Supplementary Materials Tables S5 through S9. When assessing both sample substrate types together, richness was found to not vary significantly with substrate or vegetation type. Out of the continuous variables, only pH ($\tau = 0.11$, $p < 0.05$) correlated with richness estimates. No significant differences could be retrieved for Shannon Diversity Index scores between categorical factors either, however diversity correlated negatively with carbon-to-nitrogen ratio ($\tau = -0.20$, $p < 0.001$) and positively with pH ($\tau = 0.13$, $p < 0.05$). Generalized linear models (GLMs) with diversity as the response variable gave carbon-to-nitrogen ratio (coef. = -1.5) and pH (coef. = -1.6) as significant predictors, while models for richness gave the Polar vegetation type (coef. = -496).

Following initial analyses of all samples together, substrates were subset and analysed separately. Between Kendall's rank correlation tests and GLMs, alpha diversity indices did not consistently correlate to the same variables across substrates (Figures 6 and 7; table 2). When both substrates were analysed together, significant

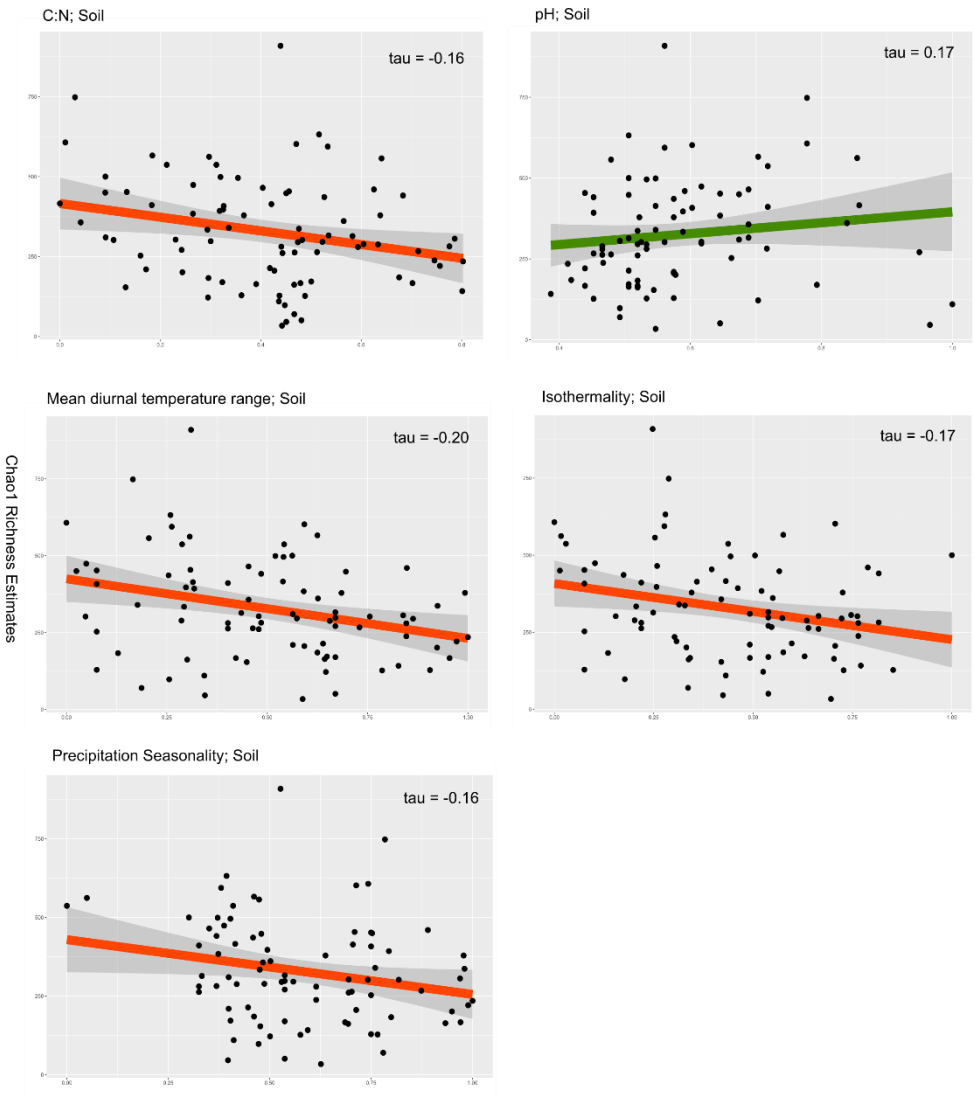

Figure 6: Significant correlations (Kendall's rank correlation tau) between Chao1 Richness Estimates (y-axis) and environmental variables; soil. Note that regression lines are illustrative for directionality, please see tau for slope indication.

correlations were majorly reflective of those observed also in soil. Vegetation type was not found to be a significant predictor of richness in either substrate and only between Broad leaf and Pine vegetation types for soil diversity (Kruskal-Wallis rank sum test, $\chi^2 = 18.228$, $p < 0.05$; Dunn's test, Bonferroni adjusted $p < 0.05$). GLMs did not indicate vegetation types as significant predictors of diversity in either substrate and for richness, only the Polar vegetation type was significant and only in soil (coef. = -726, $p < 0.05$).

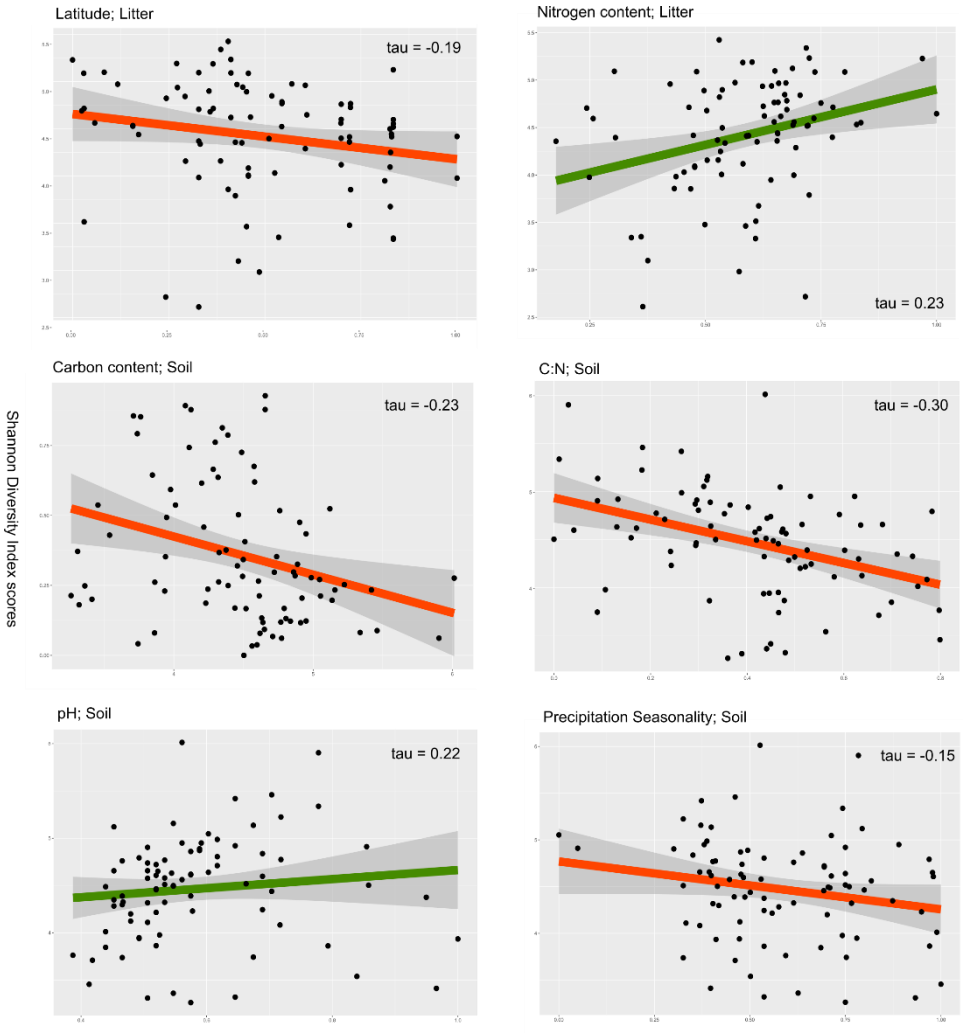

Figure 7: Significant correlations (Kendall's rank correlation tau) between Shannon Diversity index scores (y-axis) and environmental variables; Litter (top) and soil. Note that regression lines are illustrative for directionality, please see tau for slope indication.

Soil richness further correlated positively with pH ($\tau = 0.18$); negatively with carbon-to-nitrogen ratio ($\tau = -0.16$); mean diurnal temperature range ($\tau = -0.20$); isothermality ($\tau = -0.17$) and precipitation seasonality ($\tau = -0.15$) (Figure 6). Soil pH (coef. = -561), metres above sea level (coef. = -487) were indicated separately as significant soil richness predictors from GLMs. Litter richness did not correlate significantly with any continuous variables or yield significant predictors from the GLM. Litter diversity, however, correlated positively with nitrogen content ($\tau = 0.23$) and negatively with latitude ($\tau = -0.19$), and the GLM returned pH (coef. = -3.1), along with metres above sea level (coef. = 2) as significant diversity predictors (Figure 7). Soil diversity correlated negatively with carbon content ($\tau = -$

0.22), carbon-to-nitrogen ratio (τ = -0.29) and precipitation seasonality (τ = -0.15); and positively with pH (τ = 0.22). GLMs conversely returned carbon content (coef. = -2.3) and pH (coef. = -2.8) as significant diversity predictors in soil.

Table 2: Generalized linear model output for Chao1 Richness Estimates (top) and Shannon Diversity Index (bottom) – only significant predictors displayed, further details can be found in Tables S8 and S9 in Supplementary Materials.

Generalized Linear model output for ***Chao1 Richness Estimates***

Characteristic	Both substrates			Litter			Soil		
	Beta	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**
Vegetation type									
Polar	-496	-952, -41	**0.034**	14	-646, 673	>0.9	-726	-1,389, -63	**0.036**
pH	-261	-566, 45	0.1	-302	-851, 248	0.3	-561	-1,026, -97	**0.021**
Metres above sea level	-216	-535, 102	0.2	150	-319, 618	0.5	-487	-947, -27	**0.043**
Sample read abundance	0.01	0.00, 0.01	**0.007**	0.01	0.00, 0.01	**0.004**	0	0.00, 0.01	0.5
% unique OTUs in sample	16	10, 23	**<0.001**	14	1.9, 26	**0.027**	18	8.3, 27	**<0.001**
No. Obs.	168			84			84		
Null deviance	4,081,323			1,764,665			2,276,017		
Null df	167			83			83		
Deviance	2,593,928			941,145			1,038,469		
Residual df	137			54			54		

[1] CI = Confidence Interval

Generalized Linear model output for ***Shannon Diversity Index***

Characteristic	Both substrates			Litter			Soil		
	Beta	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**
C	0.31	-0.90, 1.5	0.6	-0.68	-4.0, 2.6	0.7	-2.3	-4.3, -0.30	**0.029**
C:N	-1.5	-2.9, -0.20	**0.026**	-0.1	-3.9, 3.7	>0.9	1	-1.2, 3.3	0.4
pH	-1.6	-2.7, -0.42	**0.009**	-3.1	-5.1, -1.0	**0.005**	-2.8	-4.3, -1.2	**0.001**
Metres above sea level	0.31	-0.89, 1.5	0.6	2	0.27, 3.8	**0.028**	-1.3	-2.9, 0.25	0.11
Sample read abundance	0	0.00, 0.00	**<0.001**	0	0.00, 0.00	**<0.001**	0	0.00, 0.00	>0.9
% unique OTUs in sample	0.04	0.02, 0.06	**0.001**	0.02	-0.02, 0.07	0.3	0.06	0.03, 0.09	**<0.001**
No. Obs.	168			84			84		
Null deviance	56.4			30.4			26		
Null df	167			83			83		
Deviance	37			13.1			12.1		
Residual df	137			54			54		

[1] CI = Confidence Interval

Library size (sample read abundance) and percent unique OTUs were often present as significant correlators or predictors, however as these were indicated to vary directly with richness (see earlier results on alpha diversity indices intercorrelations) they may be ignored from further inference.

3.4 Beta diversity trends in soil eukaryotes

In analysing beta diversity, NMDS ordinations with significant ($p < 0.05$) variable vectors indicate differing community structure responsivities between litter and soil (Figure 8). Ordinations also indicate grouping of community structure by vegetation type and further by

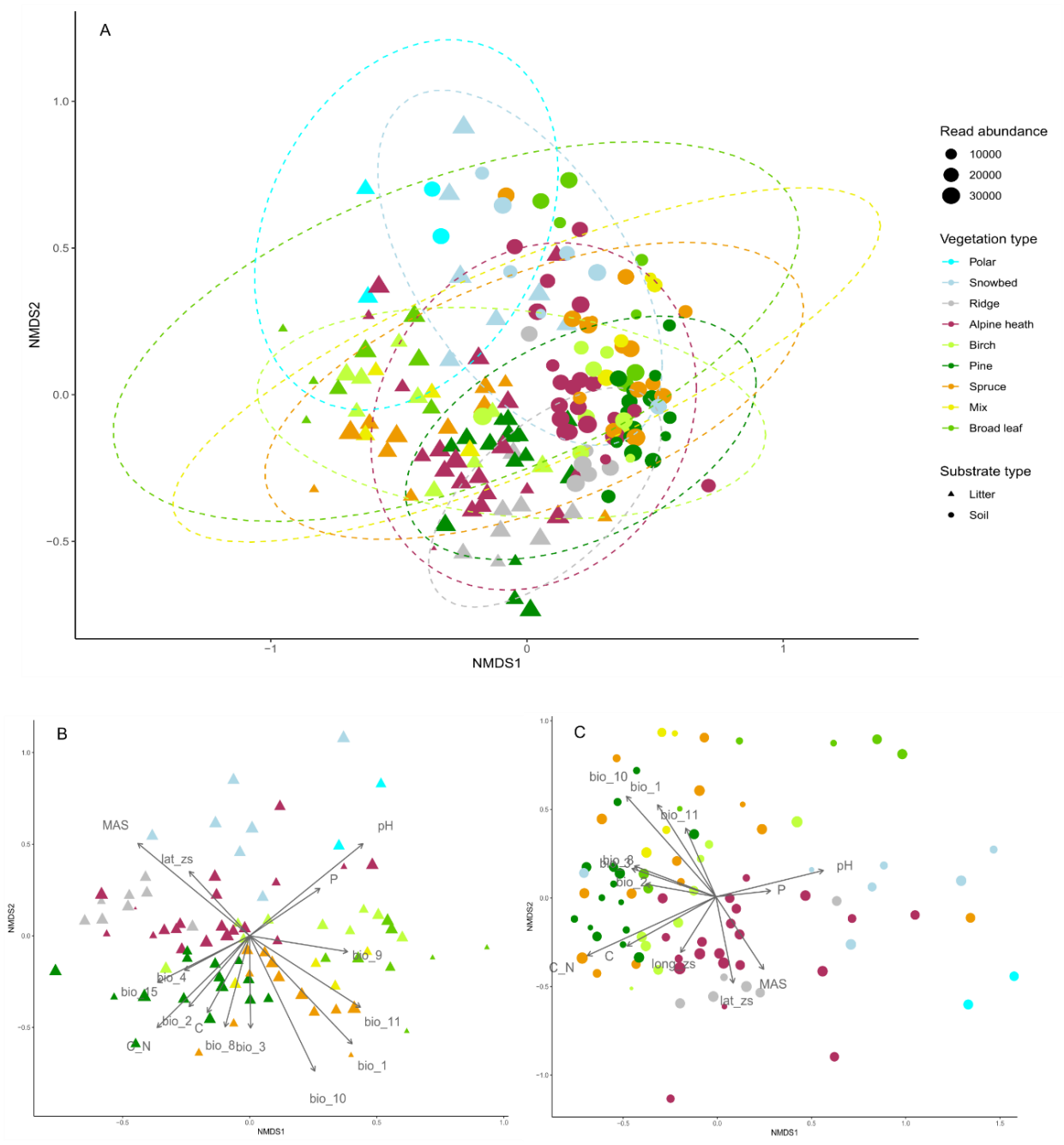

Figure 8: NMDS ordination plots for samples by vegetation types for A) Both substrates together, with t-type ellipses for vegetation types; B) Litter samples with significant ($p < 0.05$) environmental variables fitted; and C) Soil samples with significant ($p < 0.05$) environmental variables fitted.

forest and alpine vegetation types. PERMANOVAs indicate that all factors were significant to community structure when analysing both substrates together and in litter separately, while all except plant richness were significant for litter communities ($p < 0.05$) (see Table S10 in Supplementary Materials for full output summaries). It should be noted, however, that most of the significant R^2 statistics are minute (R^2 <0.03-0.08>) and so their true significance should

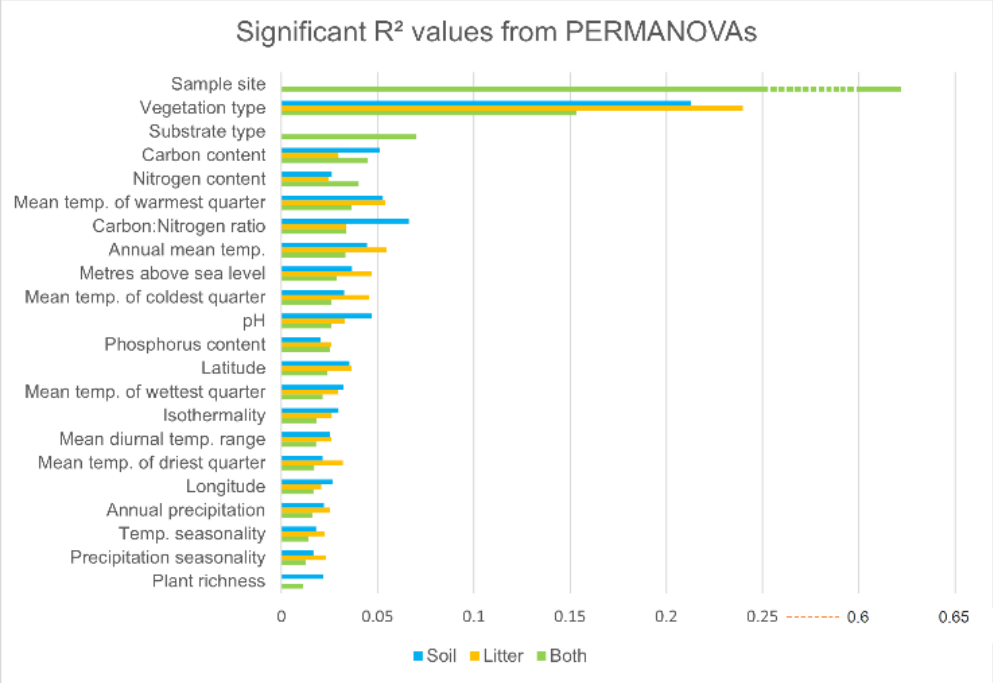

Figure 9: Significant R^2 statistics from individual-variable PERMANOVAs, note the scaling of Sample Site effects.

be interpreted with caution (Figure 9). The only variable that appears to emerge as a truly significant factor when substrates are considered separately is the dominant vegetation type (R^2 = 0.24 in litter samples; R^2 = 0.21 in soil). When viewed together, the substrate does appear to also play a considerable part in community structure (R^2 = 0.07) while sample site effects are by far the greatest contributor to overall community dissimilarity (R^2 = 0.62).

There was a significant correlation with geographic location on community structure when substrates were considered together (Mantel statistic based on Spearman's rank correlation rho, $r = 0.1185$, $p < 0.05$) as well as in litter ($r = 0.17$, $p < 0.05$) and soil ($r = 0.17$, $p < 0.05$) separately.

4. Discussion

In brief, a large diversity of OTUs is recovered but data characteristics may indicate that the observed diversity is presenting a subset of true diversity. Despite this, there are consistent variations between soil and litter in alpha diversity correlations and in community structure. The structure also varies with vegetation type, however the uncertainties associated with abundance signal limit the ecological inferences that can be made from these structural variations. Ultimately, long-read sequencing provides a solid and highly informative dataset, however some caveats must be discussed in light of the above results.

4.1 Long-read sequence data

The first and perhaps most crucial item of discussion following the results is the relationship between observed species accumulation curves and the positive richness-library size correlation. While the latter may be an indication of insufficient sequencing depth (see Figure 5), the species accumulation curves imply satisfactory sequencing in that despite variations in library size (and some severe size outliers), most sample curves still appear to reach a plateau (see Figure 2). As such, one should consider the possibility that the current dataset is comprised of only a subset of the true community diversity and that this subset is a result of community structure properties or other biological factors highlighting limitations to current eDNA metabarcoding approaches. If so, there might be an incentive to assess more deeply the suitability of long-read metabarcoding as a community mapping tool and increase knowledge on biases introduced through broad-scale taxa mapping. That being said, if this diversity selection phenomenon is real it could also be a result of unknown selective processes in pre-sequencing protocols. Investigations into whether methodological processes might be causative are ongoing as the DNA used for this study has also been used in a parallel short-read 18S-based study with separate protocols (Ella Thoen et al., in preparation) – if these samples exhibit similar plateauing it would be indicative of sample processing or DNA isolation related selection, whereas otherwise the amplification steps should be considered. There are instances of primer choice resulting in differential richness measures in metabarcoding, as recorded in common COI primers (Hajibabaei, Porter, Wright & Rudar, 2019). As this study employs primers previously used in long read sequencing of eukaryotes (Jamy et al., 2019) where such a diversity subsetting does not appear to have been an issue, one might consider primer choice less likely to be causal here (though still possible). DNA extraction protocol has also been

found to have significant impacts on recovery of community structure in soil (Dopheide et al., 2017; Santos et al., 2017). It remains, though, that a likely source of this selection bias - if it is protocol-related - is the amplification process. As the PCR approaches in this study were not designed for a methodological study, unfortunately the application of differing PCR approaches is non-random and the statistical power in the skewed distribution of final treatment groups too poor to retrospectively gain any insight on this matter. In addition to the protocols themselves, there may be factors in the samples which inhibit or limit amplification of certain groups, as overall PCR inhibition in soil- litter derived eDNA is well-documented (see review by Alaeddini, 2012) – whether these inhibitors could selectively interact with amplification processes is a more controversial question. Another aspect to consider is the availability of reagents during PCR and if the increase in amplicon length could be affecting the rate of expenditure or saturation to limit amplification. If this potential phenomenon of selective amplification is related to the amplification of long amplicons specifically, the implications for long-read sequencing in diversity mapping and particularly beta diversity are considerable.

Another interesting and amplification related find was that detected chimeric rates were very low throughout the sample data. Several approaches were tested for chimeric detection, however none yielded greater rates than the use of adjusted *uchime_denovo* which was ultimately utilized. In contrast to the 0.61% (13.27% including borderlines) rate observed in this study, Heeger et al. (2018) report a chimera formation rate of 16 % from similar PCR conditions in long-read sequence data from eDNA. Furthermore, the chimera prevalence was unexpectedly high (59-86 %) in the mock community samples in comparison to environmental samples. The main issue considered was confidence in the detection of multi-parent chimeras (or "grand-parent"). While some time was invested in testing for the most appropriate chimera detection protocol, it is still possible that this vast discrepancy between expected prevalence, mocks and collected samples is due to some volume of undetected chimeras in the collected sample data. However, the overall data resolution and phylogenetic signal retained in the study are not indicative of an excess of chimeras. And though no formal tests were conducted to assess intra-sample phylogenetic diversity, one might expect environmental samples to be fairly phylogenetically heterogenous in comparison to both mock and amplicons from narrow taxa-targeted samples. Applying broadly targeting primers might result in a lower chimeric formation rate than what could be expected from amplification of more selectively phylogenetically linked taxa. In short read amplification, community complexity is associated with increased chimera formation, however similarity between parental sequences is also an

apparent prerequisite for chimera formation, and as such the above theory may hold yet (Aas, Davey & Kauserud, 2016). There are also the differing extraction methods and concentration/compositional aspects to consider with regards to the discrepancy between mock and environmental samples.

Continuing through the long-read data curation process, a brief mention of the use of PR^2 database hits as a filtering tool is due. As an additional denoising step, sequences that could not be matched to any PR^2 database entry were excluded on the basis that they may not be eukaryote sequences. There are alternatives to this approach, such as *Barrnap*'s eukaryote detection function (https://github.com/tseemann/barrnap). Still, the conservative approach taken throughout data curation, along with the overall study aims, allowed for the more exclusive PR^2-matching strategy. However, any exclusion of sequences at this stage in data curation – that is, after all major filtering, de-noising and chimera removal – runs the risk of excluding true diversity. Especially in poorly mapped communities, "dark taxa" may be numerous and could affect community structure (Kortmann et al., 2022; Zamkovaya et al., 2021), as well as simply being diversity worth mapping on its own. For further investigations into the unmapped taxa of soil communities in this study, one approach could be to align the discarded sequences with the established reference library and work from there to establish phylogenetic placement likelihoods and hypothesize ecological functionalities.

4.2 Phylogenetic diversity and abundance signal

While diversity analyses may be considered with some care, the recovered phylogenetic topology can be said to stand apart and the observed diversity discussed. Several large groups are delimited fairly cleanly, with no major conflicts to current hypotheses of the eukaryote tree, except for perhaps the placement of Microsporidia outside of Fungi (Burki et al., 2020). It should be noted that several deep nodes are poorly supported, however this is to be expected from this level of phylogenetic signal and the use of a single DNA-locus. The branches associated with OTUs annotated as Amoebozoa, Discoba and Microsporidia are somewhat entangled and Microsporidia branches are considerably extended relative to the overall tree. In broad terms, the distribution of diversity can be described in three groups: Fungi, Cercozoans and the rest. The rest are dominated by Alveolata, then almost evenly dispersed across the remaining Stramenopiles, Amoebozoa, Discoba, Streptophyta, Chlorophyta and finally some minor straggler groups (ex. Apusozoa). Cercozoa make up another rough third of the observed

diversity and then lastly Fungi. Fungi are, when viewed together, the major diversity contributors, with Ascomycota leading the charge. Basidiomycota follow suite, however with regards to non-Dikarya/early diverging fungi the annotation and thus diversity distribution is a little less clear. The non-Dikarya branches only had 70.59 % of OTUs annotated as Fungi without being Basidiomycota or Ascomycota. Out of the remaining OTUs, approximately half were a mix of non-Fungi annotations and the other half were annotated as Basidiomycota. This discrepancy between annotation and phylogeny could be caused by an issue with the phylogenetic inference approach but could also be highlighting a database coverage issue for non-Dikarya fungi. The latter is perhaps more likely as this group holds many cryptic taxa and as-of-yet unresolved lineage relationships (Voigt et al., 2021). The annotation approach may be an important factor regardless – even if poor coverage is the issue, had the annotation been done by extracting OTUs by phylogenetic affiliation and accepting lower quality matches in databases, a closer lineage match might have taken precedence over out-of-group better matches that might be matching to regions less relevant for delimitation in the original group.

Keeping in mind that the described group delineations are approximate and not necessarily made to comparable rank levels, for further uses of this annotation approach a more rigorous and detailed topology investigation may be required. This in turn evokes another set of considerations for future phylogenetic inference in diverse organisms, such as whether the chosen clustering similarity level is reflective of true inter-taxon variation or even if the same clustering level is appropriate across groups. For instance, this issue has been raised with regards to obstacles in studying prokaryotes by applying methods and concepts from eukaryote mapping (Lara, Singer & Geisen, 2022), however issues regarding species concepts and lineage separation are relevant for untangling microeukaryote diversity as well. Increasing the volume of per-individual genetic information may be one path towards a more holistic species concept – and long-read sequencing a step on that path. In addition to some of the already discussed issues, some experimentation may also be due in other aspects of phylogenetic inference approaches. In initial alignment efforts, the high variability of certain regions and the considerable span of data even in conserved regions caused highly spaced alignments upwards of 300 000 bp. Following removal of 50% absence gaps alignment length became more manageable and pairwise match percentage was increased. Due to the hypervariability frequently observed in the ITS regions, the forced alignment of these further exacerbated good quality alignments overall. Excluding the ITS regions from alignments improved alignments and as clustering was done on full-length sequence data there was no change to OTU diversity.

This approach was successful, however the usefulness of commonly applied alignment and trimming principles and perhaps even tree-generating models might not be suited for use with long-read sequence data off the cuff, similarly to several other aspects of the long read data curation.

Moving to the relationship between observed diversity and observed abundance, some groups are disproportionately abundant relative to the observed diversity of taxa. The per-sample abundance distribution varies, however visual representations indicate that most samples generally follow a broadly similar distribution of abundance with somewhat consistent levels of separation between litter and soil (Figure 10). Despite this differentiation between litter and soil, Streptophyta is consistently and strongly represented in opposition to the low diversity observed. In the same vein, Cercozoans appear to be underrepresented in relative abundance compared to their wide phylogenetic diversity. Relative abundance is, however, *relative*, and so disproportionate shifts in one group do not indicate a likewise decrease in the remaining groups. And in addition to the previously discussed issues with a possible diversity threshold in the data and its inherent ties to abundance distributions, there is a suite of other confounding factors at play which may render the observed abundance signals unviable. For one there are the morphological scale differences in the target organisms – while some are macroscopic, complex, and come in large tissue chunks (ex. Streptophyta, Arthropoda, some Fungi), a considerable proportion is expected to be microscopic, unicellular and not necessarily clustered together. The effects of these differences can be exacerbated through behavior as larger organisms may move around more relative to their size and randomize their sample prevalence (Zinger et al., 2019). In addition, operon copy numbers have been found to be highly variable among organisms and could further inflate relative abundance prevalence (Prokopowich, Gregory & Crease, 2003). Despite homogenizing efforts, this may still result in overrepresentation of larger organisms through greater overall physical tissue and genetic presence in a sample. As read data is not controlled against individual organism contribution potential, all sequences are essentially given equal weight irrespective of how many read copies that can be expected to originate from one individual – which when viewed as relative abundance proportions can be inaccurate. This may be mediated somewhat by separating and analysing abundance data based on expected source contributions, e.g. unicellular taxa separately.

However, there are also other processes of taphonomy/environmental stochasticity to consider when working from eDNA and targeting organisms with different morphologies and ecologies.

Individual and DNA degradation may vary and interact with climate factors and collection site topography, resulting in read abundances that are not representative of the community they are collected from (see review by Harrison, Sunday & Rogers, 2019; Barnes & Turner, 2016). In addition, differential temporal turnover has been indicated for some soil organism groups (Martinović at al., 2021), which in combination with slow degradation of genetic material may misrepresent community structures (Carini et al., 2016). Soil communities also exhibit high spatial heterogeneity across small scales, both vertically and horizontally (Fiore-Donno et al., 2022; Štursová, Bárta, Šantrůčková and Baldrian, 2016). Relic DNA and intrasite heterogeneity is more relevant to the issue of separating out the active community composition and accurately placing it relative to its environment than it is to taphonomic processes but is equally important in the context of the approaches in this study.

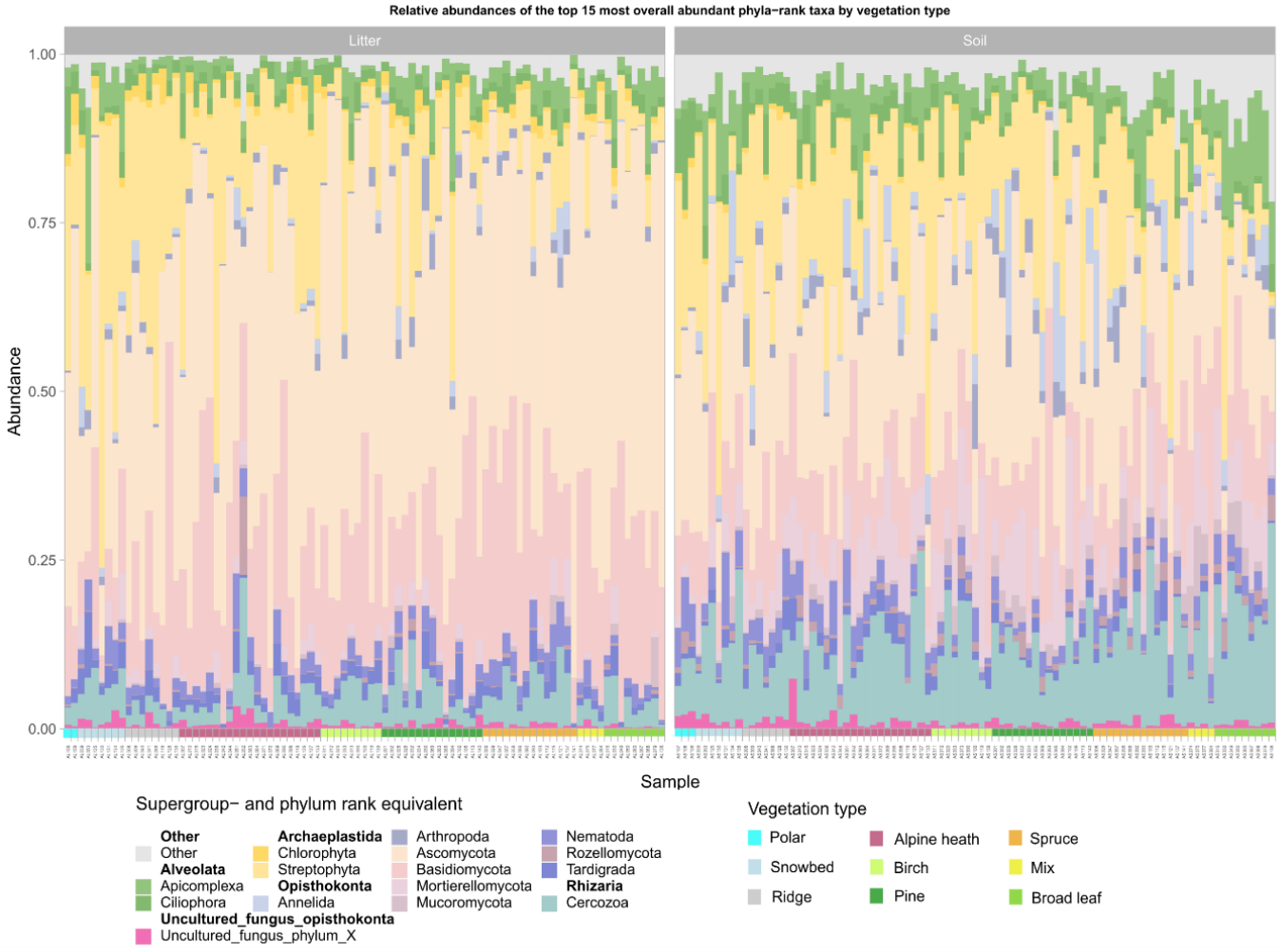

Figure 10: Per-sample relative abundance distribution based on the 15 most abundant phyla-level taxa; sorted by vegetation type along the horizontal axis and displaying substrates separately. Note that the phylum legend order within supergroups corresponds to stack order and not phylogenetic groupings.

In addition to the above issues, another and more general point may be made about the inference of community-habitat interactions when considerable chunks of the trophic chain in question are left out. In the case of this study, for instance, eukaryotes are targeted and the aim is to uncover potential trends and their relationship to the observed community composition. However, without including bacteria, archaea and other groups that contribute to trophic networks in soil, any finds and hypotheses on structural variation may in fact be describing effects on other trophic layers and the resulting cascade into the studied groups. Proportions of trophic groups have been found to shift with temperature in soil communities (Dahl et al., 2023) and the structure of trophic links themselves may also be responsive to external factors (Xiong et al., 2021). Considering that the diversity of non-eukaryotes in soil is indicated to be overwhelming relative to its eukaryote neighbors (Hudson et al., 2015), the trickle-down/trickle-up effects from these groups through trophic linkage should perhaps not be underestimated.

4.3 Diversity trends

In light of all this, including the potential issue of the data representing only a diversity subset, investigations into diversity drivers and factors shaping community structure may ultimately be describing incomplete systems. However, as some overarching trends can be hypothesized from the results, these will be discussed. The most prominent trend that can be retrieved from the data is that vegetation type is a strong predictor of community structure, both in soil and in litter. And while vegetation type is relevant to both substrates, separate NMDS plots appear to indicate that community structure in litter samples varies more homogenously across vegetation types and that, comparatively, soil samples display more intravariation in community structure between vegetation types. When plotted together, overall intervariation among soil samples becomes apparently lower than that observed for litter, however. In addition to the vegetation type groupings, the clustering of vegetation types generally align to form two supergroups: highland/alpine and lowland/forest. Keeping in mind that PERMANOVAs indicate very low explanatory power for all significant factors with the exception of vegetation type, substrate and site, the fitted NMDS vectors for significant environmental variables appear to further attenuate the highland/lowland supergrouping, at least in soil. Temperature and precipitation (climatic) factors group towards lowland/forest types while latitude, elevation and some soil edaphic factors group towards highland/alpine

types. In litter, similar trends can be observed, however overall scatter of factor directionality is greater and may indicate a more heterogeneous response in community structure. This falls somewhat in line with observed correlations in alpha diversity indices - however as only soil diversity was found to vary with vegetation type, and only significantly in one pair, the relationship between diversity, richness and community structure indicates that they are not responsive to the same processes.

Returning to PERMANOVA results, in both substrates together vegetation type was only surpassed by site effects. Site effects are a catch-all factor as they encompass all site characteristics and makes further inferences more or less redundant without highly extensive metadata records. One thing that can be extrapolated from this, however, is that inter-site variation is high. In light of the observed overlap in taxa commonality and abundance, this could align with the concept of a "core" composition of highly abundant and more generalist taxa with a smaller, more nebulous network of locally adaptive taxa (Malard et al., 2022; Barberán et al., 2012). This also coheres with the high proportion of unique OTUs found overall and the relatively low per-sample proportions. Running community structure analyses where only taxa below a global abundance threshold are included could shed some light on how dynamics shift between the less frequent taxa.

Site effects may also be inherently tied to geographical properties, and Mantel tests on both substrates found that geographic coordinate position of the sample site was significant to the structural placement of both litter and soil. As with site effects, spatial distance, too, is a rather opaque factor from which not much can be gleamed without relating to changes in elevation, latitude or longitude. However, geographic structural variation could also indicate some inherent level of spatial dispersal effect on community composition. And as Mantel r statistics were similar between substrates, this would then likely be pervasive through both litter and soil. In terms of geography in relation to vegetation, the vegetation type categories are not indicative of elevation as much as they are a product of elevation and latitude – and likely in a non-linear fashion as well if one takes into account the topography of Norway. Still, the findings correspond with the general idea that elevation is not a strictly linear predictor of richness or community structuring in terrestrial microbiomes (Wang et al., 2022; Adamczyk et al., 2019).

No formal analysis was done to determine significant taxa variation between vegetation types. However, from visualizations of relative abundance per sample, some generalized trends may

be described (Figure 10). In litter samples, relative abundances of Streptophyta are highest towards the highland/alpine vegetation types and decrease to minimal presence across the lowland/forest types. A similar trend can be seen in soil, however the decline in abundance is less pronounced and overall abundance of Streptophyta is lower. In litter, the downwards shift in Streptophyta abundances is accompanied by increasing Ascomycota and to some extent Basidiomycota prevalence, while other major groups shift more stochastically. Conversely, in soil, the moderate decline in Streptophyta corresponds with some uptick in Cercozoa, Alveolata and low-abundance taxa (categorized as *Other* in visualizations). Across soil samples, relative abundance distributions are more even among major groups compared to litter samples, with Dikarya fungi making up less of the relative abundance. From this, a small leap can be made to infer that soil communities might be more homogenous and/or stable systems compared to litter communities. Continuing with this hypothesis, this could be due to a relative isolation from environmental stochasticity through physical and chemical buffer systems such as biocrusts and small-scale chemical buffering from organisms (Weber et al., 2022; Gutiérrez & Jones, 2006). However, as microbial communities have been found to change in response to extreme geochemical and climatic fluxes (Kato et al., 2015; Stenuit & Agathos, 2015; Beyens et al., 2008), this hypothesis only holds under the assumption that the stress on the buffer systems and taxa remains stable and low. Still, these trends, along with significant indicators of differing community structures between litter and soil, come together to create a general picture of ground microbiomes as vertically stratified, at least with regards to microeukaryote taxa.

With regards to alpha diversity indices, relevant variable correlations appear to differ between litter and soil consistently, indicating that alpha diversity is shaped by different mechanisms in the two substrates. And while soil edaphic factors often correlate with alpha diversity indices, there are instances of both temperature and precipitation correlations too (Figures 6 and 7; Table 2). The general absence of strong correlations across both substrates and to vegetation types in alpha diversity indices suggests that community structure varies independently of overall richness or diversity and that neither is a very relevant factor in the mechanisms deciding community structural properties. Alternatively, the alpha diversity of functional groups in a community could be correlating strongly with certain environmental gradients or groupings – but it would depend on the group prevalence and the compositional structure of the community whether that response was observable by alpha diversity indices. Furthermore, the wide geographical scatter of collection sites presents another opportunity for differential

interactions between variables and alpha diversity – nematode diversity, for instance, has been found to correlate differentially to variables along a latitudinal gradient (Shao et al., 2023). This further follows on top of the overall correlational relationships between variables themselves (see Figures S1 and S2 in supplementary material).

That being said, some interesting correlations were still discovered. pH was perhaps the most frequently occurring correlational relationship, both in richness and diversity, but not always in both substrates. In soil richness and soil diversity, Kendall's rank correlations against pH gave a significant positive coefficient - however in GLMs these coefficient estimates were given as negative. This is an interesting discrepancy at first glance but might turn out to be an artefact of sample outliers. The GLM for soil richness also gives vegetation type Polar as a significant predictor, i.e. samples from Svalbard collection sites. As Kruskal-Wallis tests did not indicate Polar samples as significantly differing from others in richness, and the soil Polar pool is very small (n = 2), there might be aspects to GLM calculations that are inflating or skewing the significance of variables in this particular model and indicative that excluding the Svalbard samples from the GLM might have been a pertinent move. Following this line of investigation, it might also be a good idea to assess alpha diversity trends by vegetation types separately for more robust insight into factors shaping alpha diversity.

5. Concluding remarks

The study finds that long-read sequencing is a strong tool for metabarcoding, but that there is still room for streamlining molecular and bioinformatic processes to gain the most out of this approach. A vast diversity of soil eukaryotes were recovered from soil and litter samples, however alpha diversity responsivity to environmental factors remains somewhat enigmatic. Community structure appears to vary significantly and consistently with vegetation type and potentially along an eco-elevational gradient as well. Further analyses into the collinearity of vegetation types and other factors are still needed to give a full picture of this structural trend. No clear distinction could be made between the scales at which climatic factors and soil properties interact with alpha diversity and a more holistic analysis approach may be required to untangle interactive effects between variables.

References

Aas AB, Davey ML, Kauserud H. 2016. ITS all right mama: investigating the formation of chimeric sequences in the ITS2 region by DNA metabarcoding analyses of fungal mock communities of different complexities. Molecular Ecology Resources, 17: 730-741 DOI: 10.1111/1755-0998.12622

Abarenkov K, Zirk A, Piirmann T, Pöhönen R, Ivanov F, Nilsson RH, Kõljalg U. 2022. UNITE general FASTA release for Fungi. UNITE Community. DOI: 10.15156/BIO/2483911

Adamczyk M, Hagedorn F, Wipg S, Donhauser J, Vittoz P, Rixen C, Frossard A, Theurillat JP, Frey B. 2019. The Soil Microbiome of GLORIA Summits in the Swiss Alps. Frontiers in Microbiology, 10 DOI: 10.3389/fmicb.2019.01080

Aguilar-Marcelino L, Al-Ani LKT, Castañeda-Ramirez GS, Garcia-Rubio V, Ojeda-Carrasco J. 2020. Microbial technologies to enhance crop production for future needs. In: Rastegari AA, Yadav AN, Yadav N. (eds.). New and future developments in microbial biotechnology and bioengineering: trends of microbial biotechnology for sustainable agriculture and biomedicine systems: diversity and functional perspectives. Elsevier, Amsterdam, The Netherlands.

Aislabie J, Deslippe JR. 2013. Soil microbes and their contribution to soil services. In: Dymond JR. (ed.) Ecosystem services in New Zealand – conditions and trends. Manaaki Whenua Press, Lincoln, New Zealand

Alaeddini R. 2012. Forensic implications of PCR inhibition – a review. Forensic Science International: Genetics, 6: 297-305 DOI: 10.1016/j.fsigen.2011.08.006

Arroyo AS, López-Escardó D, Kim E, Ruiz-Trillo I, Najle SR. 2018. Novel diversity of deeply branching Holomycota and unicellular Holozoans revealed by metabarcoding in middle Paraná River, Argentina. Frontiers in Ecology and Evolution, 6: 99 DOI: 10.3389/fevo.2018.00099

Aslani F, Geisen S, Ning D, Tedersoo L, Bahram M. 2021. Towards revealing the global diversity and community assembly of soil eukaryotes. Ecology Letters, 25: 65-76 DOI: 10.1111/ele.13904

Bakkestuen V, Erikstad L, Halvorsen R. 2008. Step-less models for regional environmental variation in Norway. Journal of Biogeography, 35: 1906-1922 DOI: 10.1111/j.1365-2699.2008.01941.x

Barberán A, Bates ST, Casamayor EO, Fierer N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. The ISME Journal, 6: 343-351 DOI: 10.1038/ismej.2011.119

Barberán A, McGuire KL, Wolf JA, Jones FA, Wright SJ, Turner BL, Essene A, Hubbell SP, Faircloth BC, Fierer N. 2015. Relating belowground microbial composition to the taxonomic, phylogenetic, and functional trait distributions of trees in a tropical forest. Ecology Letters, 18: 1397-1405 DOI: 10.1111/ele.12536

Barnes MA, Turner CR. 2016. The ecology of environmental DNA and implications for conservation genetics. Conservation Genetics, 17: 1-17 DOI: 10.1007/s10592-015-0775-4

Bar-On YM, Phillips R, Milo R. 2018. The biomass distribution on earth. Proceedings of the National Academy of Sciences of the United States of America, 115: 6506-6511 DOI: 10.1073/pnas.1711842115

Bates ST, Clemente JC, Flores GE, Walters WA, Parfrey LW, Knight R, Fierer N. 2013. Global biogeography of highly diverse protistan communities in soil. The ISME Journal, 7: 652-659

Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A. 2022. maps: Draw Geographical Maps. R package version 3.4.1. https://CRAN.R-project.org/package=maps

Beyens L, Ledeganck P, Graae BJ, Nijs I. 2008. Are soil biota buffered against climatic extremes? An experimental test on testate amoebae in arctic tundra (Qeqertarsuaq, West Greenland). Polar Biology, 32: 453-462 DOI: 10.1007/s00300-008-0540-y

Burki F, Roger AJ, Brown MW, Simpson AGB. 2020. The new tree of Eukaryotes. Trends in Ecology & Evolution, 35: 43-55

Burki F, Sandin M, Jamy M. 2021. Diversity and ecology of protists revealed by metabarcoding. Current Biology, 31: R1267-R1280 DOI: 10.1016/j.cub.2021.07.066

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods, 13: 581-583 DOI: 10.1038/nmeth.3869

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics, 10: 421 DOI: 10.1186/1471-2105-10-421

Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nature Microbiology, 2: 16242 DOI: 10.1038/nmicrobiol.2016.242

Cavalier-Smith T, Lewis R, Chao EE, Oates B, Bass D. 2009. *Helkesimastix marina* n.sp. (Cercozoa: Sainouroidea superfam. n.) a gliding Zooflagellate of novel ultrastructure and unusual ciliary behaviour. Protist, 160: 452-479 DOI: 10.1016/j.protis.2009.03.003

Cavicchioli R, Ripple WJ, Timmis KN, Azam F, Bakken LR, Baylis M, Behrenfeld MJ, Boetius A, Boyd PW, Classen AT, Crowther TW, Danovaro R, Foreman CM, Huisman J, Hutchins DA, Jansson JK, Karl DM, Koskella B, Welch DBM, Martiny JBH, Moran MA, Orphan VJO, Reay DS, Remais JV, Rich VI, Singh BK, Stein LY, Stewart FJ, Sullivan MB, van Oppen MJH, Weaver SC, Webb EA, Webster NS. 2019. Scientists' warning to humanity: microorganisms and climate change. Nature Reviews Microbiology. 17(9): 569-586 DOI: 10.1038/s41579-019-0222-5

Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, Packer M, Crvalho GR, Blaxter ML, Lambshead PJD, Thomas WK. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Molecular Ecology, 19: 4-20 DOI: 10.1111/j.1365-294X.2009.04473.x

Crowther TW, van den Hoogen J, Mayes MA, Keiser AD, Mo L, Averill C, Maynard DS. 2019. The global soil community and its influence on biogeochemistry. Science, 365: eaav0550 DOI: 10.1126/science.aav0550

Dahl MB, Söllinger A, Sigurđsson P, Janssens I, Peñuelas J, Sigurdsson BD, Richter A, Tveit A, Urich T. 2023. Long-term warming-induced trophic downgrading in the soil microbial food web. Soil Biology and Biochemistry, 181: 109044 DOI: 10.1016/j.soilbio.2023.109044

De Mendiburu F. 2021. agricolae: Statistical procedures for Agricultural Research. R package version 1.3-5. https://CRAN.R-project.org/package=agricolae

Del Campo J, Sieracki ME, Molestina R, Keeling P, Massana R, Ruiz-Trillo I. 2014. The others: our biased perspective of eukaryotic genomes. Trends in Ecology and Evolution, 29: 252-259 DOI: 10.1016/j.tree.2014.03.006

Delgado-Baquerizo M, Maestre FT, Reich PB, Jeffries TC, Gaitan JJ, Encinar D, Berdugo M, Campbell CD, Singh BK. 2016. Microbial diversity drives multifunctionality in terrestrial ecosystems. Nature Communications, 7: 10541 DOI: 10.1038/ncomms10541

Dopheide A, Xie D, Buckley TR, Drummond AJ, Newcomb RD. 2017. Impacts of DNA extraction and PCR on DNA metabarcoding estimates of soil biodiversity. Methods in Ecology and Evolution, 10: 120-133 DOI: 10.1111/2041-210X.13086

Dohm JC, Peters P, Stralis-Pavese N, Himmelbauer H. 2020. Benchmarking of long-read correction methods. NAR Genomics and Bioinformatics, 2: 1-12 DOI: 10.1093/nargab/lqaa037

Dubey A, Malla MA, Khan F, Chowdhary K, Yadav S, Kumar A, Sharma S, Khare PK, Khan ML. 2019. Soil microbiome: a key player for conservation of soil health under changing climate. Biodiversity and Conservation, 28: 2405-2429 DOI: 10.1007/s10531-019-01760-5

Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008. Species detection using environmental DNA from water samples. Biology Letters, 4: 423-425 DOI: 10.1098/rsbl.2008.0118

Fick SE, Hijmans RJ. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37: 4302-4315 DOI: 10.1002/joc.5086

Fierer N. 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. Nature Reviews Microbiology, 15: 579-590 DOI: 10.1038/nrmicro.2017.87

Fiore-Donno AM, Human ZR, Štursová M, Mundra S, Morgado L, Kauserud H, Baldrian P, Bonkowski M. 2022. Soil compartments (bulk, soil, litter, root and rhizosphere) as main drivers of soil protistan communities distribution in forests with different nitrogen deposition. Soil Biology and Biochemistry, 168: 108628 DOI: 10.1016/j.soilbio.2022.108628

Francioli D, Lentendu G, Lewin S, Kolb S. 2021. DNA Metabarcoding for the characterization of terrestrial microbiota – pitfalls and solutions. Microorganisms, 9: 1-28 DOI: 10.3390/microorganisms9020361

Frenken T, Alacid E, Berger SA, Bourne EC, Gerphagnon M, Grossart HP, Gsell AS, Ibelings BW, Kagami M, Küpper FC, Letcher PM, Loyau A, Miki T, Nejstegaard JC, Rasconi S, Reñe A, Rorhlack T, Rojas-Jimenez K, Schmeller DS, Scholz B, Seto K, Sime-Ngando T, Sukenik A, Van de Waal DB, Van den Wyngaert S, Van Donk E, WolinskaJ, Wurzbacher C, Agha R. 2017. Integrating chytrid fungal parasites into plankton ecology: research gaps and needs. Environmental Microbiology, 19: 3802-3822 DOI: 10.1111/1462-2920.13827

Geisen S, Mitchell EAD, Adl S, Bonkowski M, Dunthorn M, Ekelund F, Fernández LD, Jousset A, Krashevska V, Singer D, Spiegel FW, Walochnik J, Lara E. 2018. Soil protists: a fertile frontier in soil biology research. FEMS Microbiology Reviews, 42: 293-323 DOI: 10.1093/femsre/fuy006

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 JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet AL, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R. 2013. The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Research, 41: D597-604 DOI: 10.1093/nar/gks1160

Gutiérrez JL, Jones CG. 2006. Physical Ecosystem Engineers as Agents of Biogeochemical Heterogeneity. BioScience, 56: 227-236 DOI: 10.1641/0006-3568(006)056[0227:PEEAAO]2.0CO;2

Hajibabaei M, Porter TM, Wright M, Rudar J. 2019. COI metabarcoding primer choice affects richness and recovery of indicator taxa in freshwater systems. PLoS ONE, 14: e0220953 DOI: 10.1371/journal.pone.0220953

Halvorsen R, Andersen T, Blom HH, Elvebakk A, Elven R, Erikstad L, Gaarder G, Moen A, Mortensen PB, Norderhaug A, Nygaard K, Thorsnes T, Ødegaard F. 2009. Naturtyper i Norge – teoretisk grunnlag, prinsipper for inndeling og definisjoner. Naturtyper i Norge versjon 1.0 Artikkel 1, 1-210

Halvorsen R, Skarpaas O, Bryn A, Bratli H, Erikstad L, Simensen T, Lieungh E. 2020. Towards a systematics of ecodiversity: The EcoSyst framework. Global Ecology and Biogeography, 29: 1887-1906 DOI: 10.1111/geb.13164

Harrison JB, Sunday JM, Rogers SM. 2019. Predicting the fate of eDNA in the environment and implications for studying biodiversity. Proceedings of the Royal Society B, 286: 20191409 DOI: 10.1098/rspb.2019.1409

Heeger F, Bourne EC, Baschien C, Yurkov A, Bunk B, Spröer C, Overmann J, Mazzoni CJ, Monaghan MT. 2018. Long-read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments. Molecular Ecology Resources, 18: 1500-1514 DOI: 10.1111/1755-0998.12937

Hijmans RJ. 2022. geosphere: Spherical Trigonometry. R package version 1.5-18. https://CRAN.R-project.org/package=geosphere

Horvath P, Halvorsen R, Stordal F, Tallaksen LM, Tang H, Bryn A. 2019. Distribution modelling of vegetation types based on area frame survey data. Applied Vegetation Science, 22: 547-560 DOI: 10.1111/avsc.12451

Hudson CM, Kirton E, Hutchinson MI, Redfern JL, Simmons B, Ackerman E, Singh S, Williams KP, Natvig DO, Powell AJ. 2015. Lignin-modifying processes in the rhizosphere of arid land grasses. Environmental Microbiology, 17: 4965-4978 DOI: 10.1111/1462-2920.13020

Iannone R, Cheng J, Schloerke B, Hughes E, Lauer A, Seo JY. 2023. gt: Easily Create Presentation-Ready Display Tables. R package version 0.9.0. https://CRAN.R-project.org/package=gt

Jamy M, Biwer C, Vaulot D, Obiol A, Jing H, Peura S, Massana R, Burki F. 2022. Global patterns and rates of habitat transitions across the eukaryotic tree of life. Nature, 6: 1458-1470 DOI: 10.1038/s41559-022-01838-4

Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, Bending G, Hilton S, Bass D, Burki F. 2019. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. Molecular Ecology Resources, 20: 429-443 DOI: 10.1111/1755-0998.13117

Jansson J. 2022. Soil viruses: Understudied agents of soil ecology. Environmental Microbiology, 25: 143-146 DOI: 10.1111/1462-2920.16258

Jansson JK, Hofmockel KS. 2019. Soil microbiomes and climate change. Nature Reviews Microbiology, 18: 35-46 DOI: 10.1038/s41579-019-0265-7

Jiao S, Chen W, Wang J, Du N, Li Q, Wei G. 2018. Soil microbiomes with distinct assemblies through vertical soil profiles drive the cycling of multiple nutrients in reforested ecosystems. Microbiome, 6 DOI: 10.1186/s40168-018-0526-0

Kato H, Mori H, Maruyama F, Toyoda A, Oshima K, Endo R, Fuchu G, Miyakoshi M, Dozono A, Ohtsubo Y, Nagata Y, Hattori M, Fujiyama A, Kurokawa K, Tsuda M. 2015. Time-series metagenomic analysis reveals robustness of soil microbiome against chemical disturbance. DNA Research, 22: 413-424 DOI: 10.1093/dnares/dsv023

Katoh K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution, 30: 772-780 DOI: 10.1093/molbev/mst010

Khomich M, Cox F, Andrew CJ, Andersen T, Kauserud H, Davey ML. 2018. Coming up short: identifying substrate and geographic biases in fungal sequence databases. Fungal Ecology, 36: 75-80 DOI: 10.1016/j.funeco.2018.08.002

Keeling PJ, del Campo J. 2017. Marine protists are not just big bacteria. Current Biology, 27: R541-R549 DOI: 10.1016/j.cub.2017.03.075

Ketzler G, Römer W, Beylich AA. 2021. The Climate of Norway. In Beylich AA (ed.*). Landscapes and landforms of Norway*, World Geomorphological Landscapes. Springer Nature Switzerland AG, Cham. DOI: 10.1007/978-3-030-52563-7_2

Kreader CA. 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Applied and Environmental Microbiology, 62: 1102-1106 DOI: https://doi.org/10.1128/aem.62.3.1102-1106.1996

Koblízková A, Dolezel J, Macas J. 1998. Influence of magnesium ion concentration and PCR amplification conditions on cross-species PCR. BioTechniques, 25: 38-42

Kortmann M, Roth N, Hilszczański J, Jaworski T, Morinière J, Seidl R, Thorn S, Müller JC. 2022. Arthropod dark taxa provide new insights into diversity responses to bark beetle infestations. Ecological Applications, 32: e2516 DOI: 10.1002/eap.2516

Lara E, Singer D, Geisen S. 2022. Discrepancies between prokaryotes and eukaryotes need to be considered in soil DNA-based studies. Environmental Microbiology, 24: 3829-3839 DOI: 10.1111/1462-2920.16019

Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research, 49: W293-W296 DOI: 10.1093/nar/gkab301

Liem M, Regensburg-Tyïnk T, Henkel C, Jansen H, Spaink H. 2021. Microbial diversity characterization of seawater in a pilot study using Oxford Nanopore Technologies long-read sequencing. BMC Research Notes, 14: 1-7 DOI: 10.1186/s13104-021-05457-3

Locey KJ, Lennon JT. 2016. Scaling laws predict global microbial diversity. Proceedings of the National Academy of Sciences of the United States of America, 113: 5970-5975 DOI: 10.1073/pnas.1521291113

López-Escardó D, Paps J, de Vargas C, Massana R, Ruiz-Trillo I, del Campo J. 2018. Metabarcoding analysis on European coastal samples reveals new molecular metazoan diversity. Scientific Reports, 8: 9106 DOI: 10.1038/s41598-018-27509-8

Malard LA, Mod HK, Guex N, Broennimann O, Yashiro E, Lara E, Mitchell EAD, Niculita-Hirzel H, Guisan A. 2022. Comparative analysis of diversity and environmental niches of soil bacterial, archaeal, fungal and protist communities reveal niche divergences along environmental gradients in the Alps. Soil Biology and Biochemistry, 169: 108674 DOI: 10.1016/j.soilbio.2022.108674

Martijn J, Lind AE, Schön ME, Spiertz I, Juzokaite L, Bunikis I, Pettersson OV, Ettema TJG. 2019. Confident phylogenetic identification of uncultured prokaryotes through long read amplicon sequencing of the 16S-ITS-23S rRNA operon. Environmental Microbiology, 21: 2485-2498 DOI: 10.1111/1462-2920.14636

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBNet.journal, 17: 10-12 DOI: 10.14806/ej.17.1.200

Martinović T, Odriozola I, Mašínová T, Bahnmann BD, Kohout P, Sedlák P, Merunková K, Větrovský T, Tomšovský M, Ovaskainen O, Baldrian P. 2021. Temporal turnover of the soil microbiome composition is guild-specific. Ecology Letters, 24: 2726-2738 DOI: 10.1111/ele.13896

McMurdie PJ, Holmes S. 2013. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE, 8: e61217 DOI: 10.1371/journal.pone.0061217

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson KH. 2008. Intraspecific ITS Variability in the Kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. Evolutionary Bioinformatics, 4: 193-201

Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K. 2018. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research, D1: D259-264 DOI: 10.1093/nar/gky1022

Ogle DH, Doll JC, Wheeler AP, Dinno A. 2023. FSA: Simple Fisheries Stock Assessment Methods. R package version 0.9.4. https://CRAN.R-project.org/package=FSA

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S,

Evangelista H, Fitzjohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J. 2022. vegan: Community Ecology Package. R package version 2.6-4. https://CRAN.R-PROJECT.org/package=vegan

Philippot L, Spor A, Hénault C, Bru D, Bizouard F, Jones CM, Sarr A, Maron PA. 2013. Loss in microbial diversity affects nitrogen cycling in soil. The ISME Journal, 7: 1609-1619

Pearman WS, Freed NE, Silander OK. 2020. Testing the advantages and disadvantages of short- and long-read eukaryotic metagenomics using simulated reads. BMC Bioinformatics, 21: 1-15 DOI: 10.1186/s12859-020-3528-4

Prasad S, Malav LC, Choudhary J, Kannojiya S, Kundu M, Kumar S, Yadav AN. 2021. Soil Microbiomes for Healthy Nutrient Recycling. In: Yadav AN et al. (eds.) Current trends in microbial biotechnology for sustainable agriculture, environmental and microbial biotechnology. Springer Nature Singapore Pte Ltd. DOI: 10.1007/978-981-15-6949-4_1

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE, 5: e9490 DOI: 10.1371/journal.pone.0009490

Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom EW, Adler PB, Bakker JD, Cleland EE, DeCrappeo NM, DeLorenze E, Hagenah N, Hautier Y, Hofmockel KS, Kirkman KP, Knops JMH, La Pierre KJ, MacDougall AS, McCulley RL, Mitchell CE, Risch AC, Schuetz M, Stevens CJ, Williams RJ, Fierer N. 2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. Ecology Letters, 18: 85-95 DOI: 10.1111/ele.12381

Procopio N, Ghignone S, Williams A, Chamberlain A, Mello A, Buckley M. 2019. Metabarcoding to investigate changes in soil microbial communities within forensic burial contexts. Forensic Science International: Genetics, 39: 73-85 DOI: 10.1016/j.fsigen.2018.12.002

Prokopowich CD, Gregory TR, Crease TJ. 2003. The correlation between rDNA copy number and genome size in eukaryotes. Genome, 46: 48-50 DOI: 10.1139/G02-103

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research, 35: 7188-7196 DOI: 10.1093/nar/gkm864

R Core Team. 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Revelle W. 2023. psych: Procedures for Psychological, Psychometric, and Personality Research. R package version 2.3.3. https://CRAN.R-project.org/package=psych

Ritter CD, Dunthorn M, Anslan S, de Lima VX, Tedersoo L, Nilsson RH, Antonelli A. 2020. Advancing biodiversity assessments with environmental DNA: Long-read technologies help reveal the drivers of Amazonian fungal diversity. Ecology and Evolution, 10: 7509-7524 DOI: 10.1002/ece3.6477

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ, 4: e2584 DOI: 10.7717/peerj.2584

Santos SS, Nunes I, Nielsen TK, Jacquiod S, Hansen LH, Winding A. 2017. Soil DNA extraction procedure influences protist 18S rRNA gene community profiling outcome. Protist, 168, 283-293 DOI: 10.1016/j.protis.2017.03.002

Schoch CL, Seifert KA, Huhndorf S, Schindel D. 2012. Nuclear ribosomal internal transcriber spacer (ITS) region as a universal DNA barcode marker for *Fungi*. Proceedings of the National Academy of Sciences of the United States of America, 109: 6241-6246 DOI: 10.1073/pnas.1117018109

Schwelm A, Berney C, Dixelius C, Bass D, Neuhauser S. 2016. The large subunit rDNA sequence of *Plasmodiophora brassicae* does not contain intra-species polymorphism. Protist, 167: 544-554 DOI: 10.1016/j.protis.2016.08.008

Semenov MV. 2021. Metabarcoding and metagenomics in soil ecology research: achievements, challenges, and prospects. Biology Bulletin Reviews, 11: 40-53 DOI: 10.1134/S2079086421010084

Sessich A, Mitter B. 2014. 21st century agriculture: integration of plant microbiomes for improved crop production and food security. Microbial Biotechnology, 8: 32-33 DOI: 10.1111/1751-7915.12180

Shao Y, Wang Z, Liu T, Kardol P, Ma C, Hu Y, Cui Y, Zhao C, Zhang W, Guo D, Fu S. 2023. Drivers of nematode diversity in forest soils across climate zones. Proceedings of the Royal Society B, 290: 20230107 DOI: 10.1098/rspb.2023.0107

Sjoberg DD, Whiting K, Curry M, Lavery JA, Larmarange J. 2021. Reproducible summary tables with the gtsummary package. The R Journal, 13: 570-580. DOI: 10.3614/RJ-2021-053

Stenuit B, Agathos SN. 2015. Deciphering microbial community robustness through synthetic ecology and molecular systems synecology. Current Opinion in Biotechnology, 33: 305-317 DOI: 10.1016/j.copbio.2015.03.012

Štursová M, Bárta J, Šantrůčková H, Baldrian P. 2016. Small-scale spatial heterogeneity of ecosystem properties, microbial community composition and microbial activities in a temperate mountain forest soil. Federation of European Microbiological Societies Microbiology Ecology, 92: fiw185 DOI: 10.1093/femsec/fiw185

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology, 21: 2045-2050 DOI: 10.1111/j.1365-294X.2012.05470.x

Tedersoo L, Albertsen M, Anslan S, Callahan B. 2021. Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. Applied And Environmental Microbiology, 87: e00626-21 DOI: 10.1128/AEM00626-21

Tedersoo L, Anslan S, Bahram M, Köljalg U, Abarenkov K. 2020. Identifying the 'unidentified' fungi: a global-scale long-read third-generation sequencing approach. Fungal Diversity, 103: 273-293 DOI: 10-1007/s13225-020-00456-4

Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Köljalg U, Kisand V, Nilsson RH, Hildebrand F, Bork P, Abarenkov K. 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. MycoKeys, 10: 1-43 DOI: 10.3897/mycokeys.10.4852

Teunisse GM. 2023. fantaxtic – Nested Bar Plots for Phyloseq Data. R package version 0.2.0. Available: https://github.com/gmteunisse/fantaxtic

Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Xu ZZ, Jiang L, Haroon MF, Kanbar J, Zhu Q, Song SJ, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, The Earth Microbiome Project Consortium. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature, 551: 457-463 DOI: 10.1038/nature24621

Tragin M, Vaulot D. 2019. Novel diversity within marine Mamiellophyceae (Chlorophyta) unveiled by metabarcoding. Scientific Reports, 9: 5190 DOI: 10.1038/s41598-019-41680-6

Valentini A, Pompanon F, Taberlet P. 2009. DNA barcoding for ecologists. Trends in Ecology & Evolution, 24: 110-117 DOI: 10.1016/j.tree.2008.09.011

Van der Heijden MGA, Bardgett RD, van Straalen NM. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecology Letters, 11: 296-310 DOI: 10.1111/j.1461-0248.2007.01139.x

Venter PC, Nitsche F, Domonell A, Heger P, Arndt H. 2017. The protistan microbiome of grassland soil: diversity in the mesoscale. Protist, 168: 546-564 DOI: 10.1016/j.protis.2017.03.005

Voigt K, James TY, Kirk PM, Santiago ALCM de A, Waldman B, Griffith GW, Fu M, Radek R, Strassert JFH, Wurzbacher C, Jerônimo GH, Simmons DR, Seto K, Gentenaki E, Hurdeal VG, Hyde KD, Nguyen TTT, Lee HB. 2021. Early-diverging fungal phyla: taxonomy, species concept, ecology, distribution, anthropogenic impact, and novel phylogenetic proposals. Fungal Diversity, 109: 59-98 DOI: 10.1007/s13225-021-00480-y

Wang J, Hu A, Meng F, Zhao W, Yang Y, Soininen J, Shen J, Zhou J. 2022. Embracing mountain microbiome and ecosystem functions under global change. New Phytologist, 234: 1987-2002 DOI: 10.1111/nph.18051

Weber B, Belnap J, Büdel B, Antoninka AJ, Barger NN, Chaudhary VB, Darrouzet-Nardi A, Eldridge DJ, Faist AM, Ferrenberg S, Havrilla CA, Huber-Sannwald E, Issa OM, Maestre FT, Reed SC, Rodriguez-Caballero E, Tucker C, Young KE, Zhang Y, Zhao Y, Zhou X, Bowker MA. 2022. What is a biocrust? A refined, contemporary definition for a broadening research community. Biological Reviews, 97: 1768-1785 DOI: 10.1111/brv.12862

Wei T, Simko V. 2021. corrplot: Visualization of a Correlation Matrix. R package version 0.92. Available: https://github.com/taiyun/corrplot

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J,

Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E. 2007. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, 35: D5-D12 DOI: 10.1093/nar/gkl1031

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, USA url: https://ggplot2.tidyverse.org

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, Françoi R, Grolemun G, Haye A, Henr L, Heste J, Kuh M, Pederse TL, Mille E, Bach SM, Müll K, Oo J, Robins D, Seid DP, Spi V, Takahas K, Vaugh D, Wil CWK, Yutani H. 2019. Welcome to the tidyverse. Journal of Open Source Software, 4: 1686 DOI: 10.21105/joss.01686

Wieder WR, Bonan GB, Allison SD. 2013. Global soil carbon projections are improved by modelling microbial processes. Nature Climate Change, 3: 909-912 DOI: 10.1038/NCLIMATE1951

Wilke CO. 2020. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1. https://CRAN.R-project.org/package=cowplot

Xiong W, Jousset A, Li R, Delgado-Baquerizo M, Bahram M, Logares R, Wilden B, de Groot GA, Amacker N, Kowalchuk GA, Shen Q, Geisen S. 2021. A global overview of the trophic structure within microbiomes across ecosystems. Environment International, 151: 106438 DOI: 10.1016/j.envint.2021.106438

Zamkovaya T, Foster JS, de Crécy-Lagard V, Conesa A. 2021. A network approach to elucidate and prioritize microbial dark matter in microbial communities. The ISME Journal, 15: 228-244 DOI: 10.1038/s41396-020-00777-x

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. Molecular Ecology, 28: 528-543 DOI: 10.1111/mec.14919

Zhang Y, Dong S, Gao Q, Ganjurjav H, Wang X, Geng W. 2019. "Rare biosphere" plays important roles in regulating soil available nitrogen and plant biomass in alpine grassland ecosystems under climate changes. Agriculture, Ecosystems and Environment, 279: 187-193 DOI: 10.1016/j.agee.2018.11.025

Økland RH, Økland T, Rydgren K. 2001. Vegetation-environment relationships of boreal spruce swamp forests in Østmarka Nature Reserve, SE Norway. Sommerfeltlia, 29: 1-1 DOI: 10.2478/som-2001-0001

Supplementary Information

The sequencing service was provided by the Norwegian Sequencing Centre (www.sequencing.uio.no), a national technology platform hosted by the University of Oslo and supported by the "Functional Genomics" and "Infrastructure" programs of the Research Council of Norway and the Southeastern Regional Health Authorities.

Supplementary Materials

1. PCR details

Table S1: Overview of reagent composition variations applied to samples for PCR.

Sample ID	BSA (µL)	$MgCl_2$ (µL)	Sample dilution
AL001	2	0	20x
AL002	2	0	20x
AL002 REP	2	0	1x
AL006	2	0	20x
AL007	2	0	1x
AL008	2	0	20x
AL009	2	0	20x
AL010	2	0	20x
AL011	2	0	20x
AL012	2	0	200x
AL015	2	0	1x
AL016	2	0	20x
AL020	2	0	20x
AL023	2	0	20x
AL024	2	0	20x
AL026	2	0	1x
AL028	2	0	1x
AL029	2	0	20x
AL030	2	0	20x
AL032	2	0	20x
AL033	2	0	20x
AL034	2	0	20x
AL036	3	1	20x
AL039	2	0	20x
AL040	2	0	20x
AL041	2	0	1x
AL042	1	0	1x
AL044	3	1	20x
AL047	2	0	20x
AL049	1	0	1x
AL050	2	0	20x
AL051	2	0	1x
AL052	2	0	20x
AL053	2	0	20x
AL056	2	0	20x
AL057	3	1	20x
AL058	2	0	20x
AL060	2	0	20x
AL063	2	0	20x
AL064	2	0	20x
AL067	2	0	1x
AL068	2	0	20x
AL069	2	0	20x
AL071	3	1	20x
AL072	2	0	200x
AL073	2	0	20x
AL074	2	0	20x
AL075	2	0	20x
AL077	2	0	20x
AL079	2	0	20x
AL083	2	0	20x
AL084	2	0	1x
AL085	1	0	1x
AL085 REP	2	0	1x
AL086	2	0	20x
AL088	2	0	20x
AL088 REP	2	0	20x
AL089	2	0	20x
AL090	2	0	20x
AL092	2	0	20x
AL093	2	0	20x
AL094	2	0	1x
AL096	2	0	20x
AL098	2	0	20x
AL100	2	0	20x
AL102	2	0	20x
AL103	2	0	20x
AL105	3	1	20x
AL107	2	0	20x
AL108	2	0	20x
AL109	2	0	20x
AL112	2	0	20x
AL113	2	0	1x
AL115	2	0	20x
AL116	2	0	1x
AL118	2	0	20x
AL119	2	0	20x
AL121	2	0	200x
AL125	2	0	20x
AL126	2	0	20x
AL127	3	1	20x
AL128	2	0	1x
AL130	2	0	20x
AL131	2	0	20x
AL133	2	0	20x
AL134	2	0	20x
AL135	2	0	20x
AL136	2	0	20x
AL137	2	0	20x
AL138	1	0	1x
AL139	2	0	20x
AL139 REP	2	0	20x
AL140	2	0	1x
AL141	3	1	20x
NEG1	2	0	20x
NEG2	2	0	20x

Sample ID	BSA (µL)	$MgCl_2$ (µL)	Sample dilution
AS001	2	0	20x
AS002	3	1	20x
AS006	2	0	1x
AS007	3	1	200x
AS008	3	1	20x
AS009	2	0	20x
AS010	2	0	20x
AS011	3	1	20x
AS012	1	0	1x
AS015	2	0	20x
AS016	2	0	1x
AS020	2	0	20x
AS023	2	0	20x
AS023 REP	3	1	20x
AS024	2	0	20x
AS026	2	0	20x
AS028	2	0	20x
AS029	2	0	20x
AS030	3	1	20x
AS032	2	0	1x
AS033	3	1	20x
AS034	1	0	1x
AS036	3	1	20x
AS039	2	0	20x
AS040	2	0	20x
AS041	1	0	1x
AS042	2	0	20x
AS044	1	0	1x
AS047	2	0	20x
AS049	2	0	20x
AS050	2	0	20x
AS051	2	0	1x
AS052	2	0	1x
AS053	1	0	1x
AS056	2	0	1x
AS057	2	0	20x
AS058	2	0	20x
AS060	2	0	20x
AS060 REP	2	0	20x
AS063	2	0	1x
AS064	2	0	20x
AS064 REP	2	0	20x
AS067	2	0	1x
AS068	2	0	20x
AS069	3	1	20x
AS071	2	0	1x
AS072	1	0	1x
AS073	1	0	1x
AS074	2	0	1x
AS075	2	0	20x
AS079	2	0	20x
AS083	1	0	1x
AS084	1	0	1x
AS085	3	1	20x
AS086	2	0	20x
AS088	2	0	20x
AS089	3	1	20x
AS090	2	0	20x
AS092	2	0	20x
AS093	3	1	20x
AS094	3	1	20x
AS096	2	0	20x
AS098	1	0	1x
AS100	2	0	1x
AS100 REP	2	0	1x
AS102	3	1	20x
AS103	2	0	20x
AS105	1	0	1x
AS107	2	0	1x
AS108	1	0	1x
AS109	2	0	20x
AS112	1	0	1x
AS113	2	0	20x
AS115	3	1	20x
AS116	3	1	20x
AS118	3	1	20x
AS119	3	1	20x
AS121	2	0	20x
AS125	2	0	20x
AS126	2	0	20x
AS127	2	0	20x
AS128	2	0	20x
AS130	2	0	1x
AS131	1	0	1x
AS133	1	0	1x
AS134	1	0	1x
AS135	2	0	1x
AS135	2	0	1x
AS136	2	0	1x
AS137	2	0	20x
AS138	2	0	20x
AS139	2	0	200x
AS140	3	1	20x
AS141	2	0	20x
MOCK1	2	0	1x
MOCK2	2	0	1x

Table S2: Compositional template for PCR master mixes and primers

Component	µL	Notes
Phusion Plus Buffer	4x	
BSA	2x*	*Varied between 1-3x, see details in table S1
dNTP	2x	
Phusion Plus Polyme	0.2x	
Nuclease free water	5.8x*	*Adjustments made relative to changing volumes of BSA, MgCl2 and DNA template
($MgCl_2$)	1x*	*Varied between 0-2x, see details in table S1
Forward primer	1	
Reverse primer	1	
DNA template	2*	*Some tests were done using 4 µL but none of these products were sequenced

Table S3: Agarose gel set-up for checking quantity and size of amplicon products after PCR.

Agarose gel set-up for checking PCR products

Gel set-up

75-150 ml 1% Agarose gel
3-6 µL GelRed

Electrophoresis set-up

90V
400mA
15-20 minutes

Well loading

1-0.55 µL loading buffer
3 µL PCR product

Ladder

1-0.55 µL High Range Ladder

2. Metadata statistics

Table S4: Summary of metadata variable statistics, including raw and transformed ranges and the transformation formulas applied.

Variable		Statistics for untransformed data				Transformation and *c* used		Statistics for ranged and/or transformed data			
Description	*Shortname*	*MIN*	*MAX*	*AVG*	*STD*	*TR*	*c*	*MIN*	*MAX*	*AVG*	*STD*
% Nitrogen i sample	N	0.101487	2.635571	1.026825	0.525283	ln(c+x)	2.310063	0	1	0.435058	0.219136
Carbon-to-Nitrogen ratio, calculated from raw percentages	C	1.445062	54.46666	31.97459	16.87653			0	1	0.578303	0.317428
% Carbon in sample	C_N	13.07091	75.17652	31.08951	9.962263	ln(c+x)	3.911816	0	1	0.445229	0.180163
% Phosphorus in sample	P	0.014989	0.191074	0.052473	0.023318	ln(c+x)	0.006532	0	1	0.422891	0.170283
pH measured in sample	pH	0	7.51	4.617809	0.895678	e^cx	0.057316	0	1	0.566419	0.123179
Latitude	lat_zs	58.09573	78.17125	62.82104	4.425323	ln(c+x)	-57.1989	0	1	0.491481	0.244671
Longitude	long_zs	5.908917	25.48348	11.38576	5.013666	ln(c+x)	-4.97416	0	1	0.534406	0.243995
Metres above sea level	MAS	35	1454	530.7472	390.3981	ln(c+x)	279.5669	0	1	0.484826	0.289882
Annual mean temperature	bio_1	-6.36667	7.65	2.213998	3.173236	e^cx	0.056975	0	1	0.537096	0.237495
Mean diurnal ranges (mean of monthly (max temp - min temp))	bio_2	3.875	9.558333	6.499064	1.457351	ln(c+x)	8.635897	0	1	0.496127	0.257065
Isothermality (BIO2/BIO7)(x100)	bio_3	20.07772	30.53064	25.14565	2.495762	e^c(e^cx)	0.039962	0	1	0.440151	0.23387
Temperature Seasonality (standard deviation x 100)	bio_4	544.0212	1075.771	702.7652	113.2838	ln(c+x)	-475.324	0	1	0.500245	0.220064
Max temperature of Warmest Month	bio_5*	7.3	21.6	16.40843	3.066492	e^cx	0.05087	0	1	0.56862	0.228099
Min temperature of Coldest Month	bio_6*	-22	-1.1	-9.29551	4.75569	e^cx	0.089984	0	1	0.431026	0.219953
Temperature Annual Range (BIO5-BIO6)	bio_7*	18.6	40.2	25.70393	4.69757	ln(c+x)	-14.1271	0	1	0.49596	0.222999
Mean Temperature of Wettest Quarter	bio_8	-2.1	12.1	4.319757	4.18516	ln(c+x)	56.21508	0	1	0.470902	0.296841
Mean Temperature of Driest Quarter	bio_9	-13.4333	9.516666	1.043165	5.395435	e^cx	0.044029	0	1	0.539983	0.251395
Mean Temperature of Warmest Quarter	bio_10	3.783333	15.95	11.33137	2.703981			0	1	0.61883	0.2219
Mean Temperature of Coldest Quarter	bio_11	-15.1	1.383333	-5.79831	3.898326	e^cx	0.081041	0	1	0.436965	0.22713
Annual Precipitation	bio_12	319	3085	1165.691	607.3284	ln(c+x)	131.6664	0	1	0.483573	0.237705
Precipitation of Wettest Month	bio_13*	35	397	143.1292	73.79855	ln(c+x)	-5.72507	0	1	0.544354	0.201718
Precipitation of Driest Month	bio_14*	16	116	53.30337	25.61025	ln(c+x)	51.69856	0	1	0.451274	0.269465
Precipitation Seasonality (coefficient of Variation)	bio_15	18.87791	51.12238	31.30995	7.701365	ln(c+x)	-13.8881	0	1	0.576445	0.214101
Precipitation of Wettest Quarter	bio_16*	97	1088	396.9551	209.9267	ln(c+x)	-17.1092	0	1	0.54521	0.208497
Precipitation of Driest Quarter	bio_17*	52	407	181.5955	83.55379	ln(c+x)	302.8484	0	1	0.427712	0.248686
Precipitation of Warmest Quarter	bio_18*	89	523	271.2416	81.06148	ln(c+x)	176.6092	0	1	0.522953	0.184878
Precipitation of Coldes Quarter	bio_19*	55	1002	310.1798	214.1313	ln(c+x)	34.13569	0	1	0.472315	0.258003
Plant richness per site	plant_richness	1	34	13.12921	5.234624	ln(c+x)	7.063033	0	1	0.544703	0.153738
Latitude (retained as untransformed coordinates)	lat	58.09573	78.17125	62.82104	4.425323						
Longitude (retained as untransformed coordinates)	long	5.908917	25.48348	11.38576	5.013666						
Sample read abundance	lib_size	2477	172590	20745.18	18430.56						
% unique OTUs in sample	pc_unq_otus	0	34.80903	8.729241	5.52075						

* Variables excluded from final diversity analyses

3. Metadata correlation visualizations

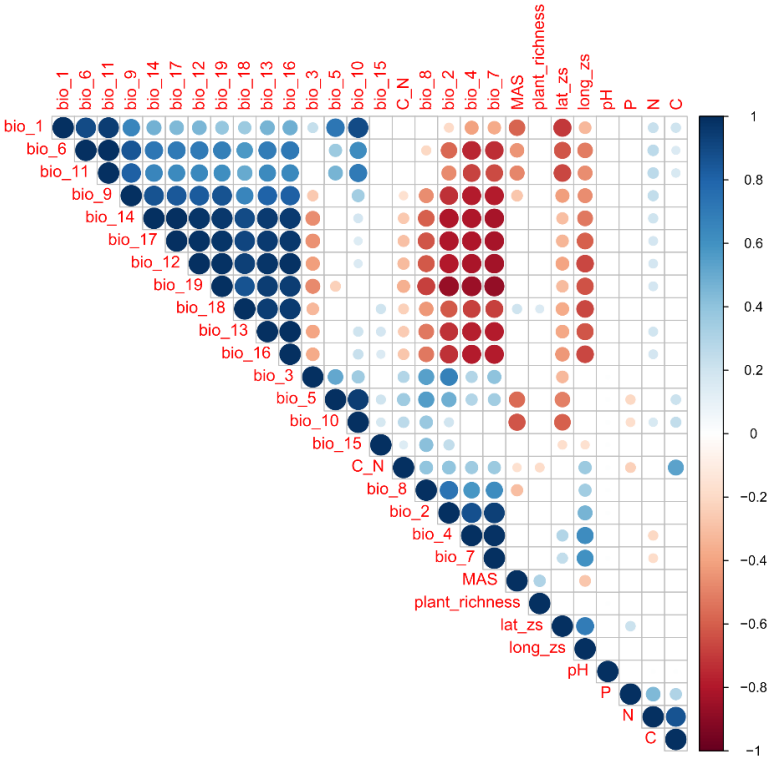

Figure S1: Correlation plot for all continuous environmental metadata variables as given by *corr.test()* from *psych* (Revelle, 2023) and default settings and visualizing with *corrplot()* from *corrplot* (Wei & Simko, 2021); showing significant correlations ($p < 0.05$) and giving the correlation coefficient's directionality by color. Short names are explained in Table S4.

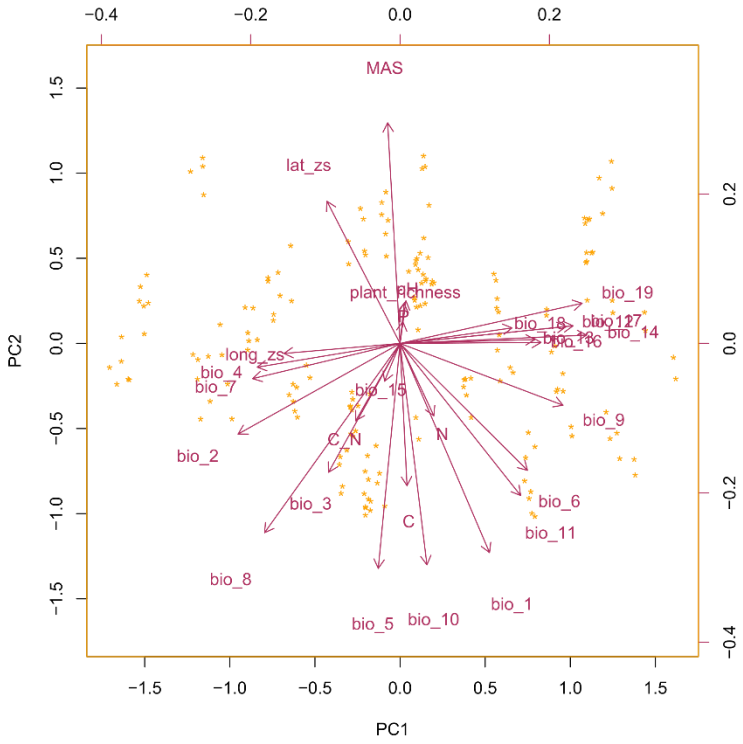

Figure S2: Principal Component Analysis visualization for interrelations between all continuous environmental metadata variables, produced by *prcomp()* from the *stats* base package. Short names are explained in Table S4.

4. Alpha Diversity statistical analysis summaries (Kendall's rank correlations and GLMs)

Table S5: Summary table for Kendall's rank correlation tests on alpha diversity indices amongst themselves and in relation to data characteristics (Sample read abundance; % unique OTUs in sample).

	Kendall's rank correlation output, outliers excluded									*Kendall's rank correlation output, outliers included*		
	Both			Litter			Soil			Both		
Character	z	τ	p-value	z	τ	p-value	z	τ	p-value	z	τ	p-value
Shannon Diversity Index : Evenness scores	8.1516	0.4210929	**3.59E-16**	7.219	0.5327731	**5.24E-13**	4.3952	0.3243697	**1.11E-05**	7.9821	0.4027804	**1.44E-15**
Shannon Diversity Index : Chao1 Richness estimates	9.9778	0.5158325	**2.20E-16**	6.9879	0.5161926	**2.79E-12**	7.2496	0.535314	**4.18E-13**	10.181	0.5141352	**2.20E-16**
Evenness scores : Chao1 Richness estimates	-1.229	-0.06353853	0.2191	0.65666	0.048507	0.5114	-1.9054	-0.1406951	0.05673	-1.6543	-0.08353982	0.09806
Evenness scores : Sample read abundance	0.18866	0.0097462	0.8504	1.5334	0.1131653	0.1252	-0.73254	-0.0540692	0.4638	-0.088061	-0.004443739	0.9298
Chao1 Richness estimates : Sample read abundance	3.0632	0.1583642	**0.00219**	2.5621	0.1892613	**0.0104**	1.2108	0.0894184	0.226	4.0761	0.2058385	**4.58E-05**
Shannon Diversity Index : Sample read abundance	2.617	0.1351944	**0.00887**	2.7783	0.205042	**0.005464**	0.80845	0.0596722	0.4188	3.2457	0.1637835	**0.001172**
Evenness scores : % unique OTUs	-0.4218	-0.02179362	0.6732	-0.77809	-0.05744711	0.4365	0	0	1	-0.57995	-0.02926984	0.5619
Chao1 Richness estimates : % unique OTUs	6.6829	0.3455629	**2.34E-11**	4.1754	0.3085554	**2.98E-05**	5.4353	0.4013454	**5.47E-08**	5.0761	0.2561905	**3.85E-07**
Shannon Diversity Index : % unique OTUs	4.7045	0.243072	**2.55E-06**	2.0762	0.1532857	**0.03788**	4.6001	0.3394958	**4.22E-06**	7.2212	0.3647224	**5.15E-13**
Sample read abundance : % unique OTUs	1.9271	0.0995718	0.05397	2.0762	0.1532857	**0.03788**	0.80085	0.0591119	0.4232	2.5563	0.12902	**0.01058**

p < 0.05

Table S6: Summary table for Kendall's rank correlation tests with Shannon Diversity Index scores and continuous environmental variables.

	Kendall's rank correlation output for Shannon-Wiener Diversity Index								
	Both			Litter			Soil		
Character	z	τ	p-value	z	τ	p-value	z	τ	p-value
Carbon content	-0.62933	-0.03250957	0.5291	1.3132	0.0969188	0.1891	-3.0516	-0.2252101	**0.002276**
Nitrogen content	0.94736	0.0489384	0.3435	3.1275	0.2308123	**0.001763**	-1.207	-0.08907563	0.2274
Carbon:Nitrogen ratio	-3.9094	-0.2019492	**9.25E-05**	-1.6093	-0.1187675	0.1076	-4.0004	-0.2952381	**6.32E-05**
Phosphorus content	1.1037	0.0570136	0.2697	0.334	0.0246499	0.7384	1.0855	0.080112	0.2777
pH	2.3645	0.1253335	**0.01805**	0.27131	0.0206887	0.7862	2.8736	0.2161032	**0.004059**
Metres above sea level	-0.1469	-0.007614437	0.8832	-0.23913	-0.01765943	0.811	0.064526	0.0047652	0.9486
Longitude	-1.6051	-0.08314169	0.1085	-1.6776	-0.1238095	0.09342	-0.57691	-0.04257703	0.564
Latitude	-1.4191	-0.07351843	0.1559	-2.6151	-0.1930242	**0.00892**	0.86158	0.0635943	0.3889
Plant richness	1.9403	0.1034493	0.05234	1.1693	0.0887865	0.2423	1.5695	0.1192066	0.1165
Annual mean temp.	0.74545	0.038748	0.456	1.8717	0.1386204	0.06125	-0.89597	-0.06634849	0.3703
Mean diurnal temp. range	-0.61876	-0.03217745	0.5361	0.45563	0.0337747	0.6487	-1.3288	-0.09839818	0.1839
Isothermality	0.66456	0.034539	0.5063	1.549	0.1147042	0.1214	-0.68716	-0.05087874	0.492
Temp. seasonality	-0.86676	-0.04504785	0.3861	-0.34168	-0.02530239	0.7326	-0.8542	-0.06324705	0.393
Mean temp. of wettest quarter	-0.81556	-0.04241254	0.4148	0.25817	0.0191282	0.7963	-1.3896	-0.1029545	0.1647
Mean temp. of driest quarter	1.0555	0.05486	0.2912	1.0782	0.0798431	0.2809	0.41002	0.0303629	0.6818
Mean temp. of warmest quarter	-0.0094363	0.0004908623	0.9925	1.6781	0.1243682	0.09332	-1.6326	-0.1209917	0.1026
Mean temp. of coldest quarter	1.0744	0.0559198	0.2826	1.8073	0.1340107	0.07072	-0.37209	-0.02759044	0.7098
Annual precipitation	0.64037	0.0333203	0.5219	0.54296	0.0402538	0.5872	0.33413	0.0247681	0.7383
Precipitation seasonality	-1.7834	-0.09268789	0.07452	-0.56947	-0.04217065	0.569	-1.978	-0.1464521	**0.04793**

p < 0.05

Table S7: Summary table for Kendall's rank correlation tests with Chao1 Richness estimates and continuous environmental variables.

Kendall's rank correlation output for Chao1 Richness estimates

	Both			Litter			Soil		
Character	z	τ	p-value	z	τ	p-value	z	τ	p-value
Carbon content	0.53096	0.0274498	0.5954	-0.28468	-0.02102903	0.7759	-1.7763	-0.1311659	0.07568
Nitrogen content	1.3368	0.0691121	0.1813	0.7933	0.0586009	0.4276	-0.75153	-0.05549328	0.4523
Carbon:Nitrogen ratio	-1.8247	-0.09433242	0.06805	-1.2412	-0.09168657	0.2145	-2.1635	-0.1597534	**0.0305**
Phosphorus content	1.6899	0.0873655	0.09104	0.42891	0.0316837	0.668	0.97927	0.0723094	0.3274
pH	2.0568	0.1091061	**0.0397**	0.70156	0.0535488	0.483	2.367	0.1780782	**0.01793**
Metres above sea level	-0.43129	-0.02237223	0.6663	0.22775	0.016835	0.8198	-0.70981	-0.05244706	0.4778
Longitude	-0.33424	-0.01732633	0.7382	-0.8844	-0.06533019	0.3765	0.083503	0.0061659	0.9335
Latitude	0.40162	0.0208225	0.688	-0.98689	-0.07291085	0.3237	1.4158	0.104555	0.1568
Plant richness	0.87891	0.0468961	0.3795	1.0665	0.0810574	0.2862	0.33524	0.0254759	0.7374
Annual mean temp.	-0.42867	-0.02229972	0.6682	-0.098715	-0.007317789	0.9214	-0.29613	-0.02194103	0.7671
Mean diurnal temp. range	-1.1486	-0.05977594	0.2507	1.2417	0.0921263	0.2144	-2.7032	-0.2002823	**0.006868**
Isothermality	-1.1566	-0.06015869	0.2474	0.73276	0.0543128	0.4637	-2.3045	-0.1707222	**0.02119**
Temp. seasonality	-0.43946	-0.0228575	0.6603	1.3099	0.0970877	0.1902	-1.8413	-0.136409	0.06557
Mean temp. of wettest quarter	-0.88163	-0.04588435	0.378	0.49738	0.036886	0.6189	-1.7161	-0.1272173	0.08614
Mean temp. of driest quarter	0.49607	0.0258042	0.6198	-0.86185	-0.06388091	0.3888	1.6249	0.1203944	0.1042
Mean temp. of warmest quarter	-0.93017	-0.04842377	0.3523	0.36829	0.0273203	0.7127	-1.4655	-0.108672	0.1428
Mean temp. of coldest quarter	-0.035051	-0.001825696	0.972	-0.42147	-0.03128108	0.6734	0.41006	0.0304228	0.6818
Annual precipitation	0.40715	0.0212017	0.6839	-0.56957	-0.04226576	0.569	1.1543	0.0856105	0.2484
Precipitation seasonality	-1.6931	-0.08806447	0.09043	-0.1177	-0.008723825	0.9063	-2.0919	-0.1549718	**0.03645**

p < 0.05

Table S8: Summary table for Generalized linear models with Shannon Diversity Index scores as response variable against all environmental and data characteristic variables.

Generalized Linear model output for Shannon Diversity Index

Characteristic	Both substrates			Litter			Soil		
	Beta	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**
Substrate									
Litter	—	—							
Soil	-0.08	-0.35, 0.19	0.5						
Vegetation type									
Alpine heath	—	—		—	—		—	—	
Birch	0.18	-0.20, 0.56	0.4	0.4	-0.13, 0.93	0.14	0.17	-0.34, 0.69	0.5
Broad leaf	-0.04	-0.73, 0.65	>0.9	0.16	-0.77, 1.1	0.7	-0.16	-1.2, 0.83	0.7
Mix	0.21	-0.44, 0.87	0.5	0.77	-0.17, 1.7	0.12	-0.09	-0.91, 0.74	0.8
Pine	0.24	-0.29, 0.76	0.4	0.54	-0.17, 1.3	0.14	-0.25	-0.96, 0.45	0.5
Polar	-0.28	-2.0, 1.4	0.8	1.9	-0.58, 4.3	0.14	-1.1	-3.3, 1.2	0.4
Ridge	-0.19	-0.55, 0.17	0.3	-0.29	-0.77, 0.19	0.2	0.14	-0.36, 0.64	0.6
Snowbed	-0.14	-0.54, 0.26	0.5	-0.33	-0.91, 0.25	0.3	0.37	-0.20, 0.94	0.2
Spruce	0.32	-0.22, 0.85	0.2	0.58	-0.15, 1.3	0.13	0.12	-0.57, 0.81	0.7
N	-0.6	-2.1, 0.90	0.4	1.5	-2.7, 5.6	0.5	1.4	-0.94, 3.8	0.2
C	0.31	-0.90, 1.5	0.6	-0.68	-4.0, 2.6	0.7	-2.3	-4.3, -0.30	**0.029**
C:N	-1.5	-2.9, -0.20	**0.026**	-0.1	-3.9, 3.7	>0.9	1	-1.2, 3.3	0.4
P	-0.07	-0.79, 0.66	0.9	0.48	-0.75, 1.7	0.4	0.02	-0.88, 0.92	>0.9
pH	-1.6	-2.7, -0.42	**0.009**	-3.1	-5.1, -1.0	**0.005**	-2.8	-4.3, -1.2	**0.001**
Latitude	0.4	-1.1, 1.9	0.6	0.9	-1.3, 3.1	0.4	0.09	-1.9, 2.0	>0.9
Longitude	-0.55	-1.8, 0.70	0.4	-0.5	-2.2, 1.2	0.6	-0.27	-1.9, 1.4	0.8
Metres above sea level	0.31	-0.89, 1.5	0.6	2	0.27, 3.8	**0.028**	-1.3	-2.9, 0.25	0.11
Annual mean temperature	-5.2	-16, 5.8	0.4	-1.9	-18, 14	0.8	-4.6	-19, 9.8	0.5
Mean diurnal temp. range	-1.5	-6.2, 3.1	0.5	4	-2.5, 10	0.2	-3.6	-9.6, 2.5	0.3
Isothermality	0.86	-1.6, 3.3	0.5	-2.3	-5.6, 1.1	0.2	3.1	-0.14, 6.3	0.066
Temperature seasonality	1.4	-2.2, 5.1	0.4	-0.56	-5.7, 4.6	0.8	2.5	-2.3, 7.2	0.3
Mean temp. of Wettest Quarter	0.18	-0.58, 0.94	0.6	-0.36	-1.4, 0.71	0.5	-0.22	-1.3, 0.82	0.7
Mean temp. of Driest Quarter	0.59	-0.64, 1.8	0.3	0.56	-1.2, 2.3	0.5	-0.26	-1.9, 1.4	0.8
Mean temp. of Warmest Quarter	2.7	-4.5, 9.9	0.5	0.02	-11, 11	>0.9	1.7	-8.1, 11	0.7
Mean temp. of Coldest Quarter	3.5	-4.0, 11	0.4	5.7	-4.6, 16	0.3	1.8	-8.1, 12	0.7
Annual precipitation	-0.69	-2.3, 0.94	0.4	-1	-3.3, 1.2	0.4	0.86	-1.5, 3.2	0.5
Precipitation seasonality	-0.39	-1.3, 0.55	0.4	-0.52	-1.9, 0.83	0.5	-0.44	-1.7, 0.78	0.5
Plant richness	0.26	-0.40, 0.91	0.4	0.68	-0.22, 1.6	0.14	0.37	-0.57, 1.3	0.4
Sample read abundance	0	0.00, 0.00	**<0.001**	0	0.00, 0.00	**<0.001**	0	0.00, 0.00	>0.9
% unique OTUs in sample	0.04	0.02, 0.06	**0.001**	0.02	-0.02, 0.07	0.3	0.06	0.03, 0.09	**<0.001**
No. Obs.	168			84			84		
Null deviance	56.4			30.4			26		
Null df	167			83			83		
Deviance	37			13.1			12.1		
Residual df	137			54			54		

[1] CI = Confidence Interval

p < 0.05

Table S9: Summary table for Generalized linear models with Chao1 Richness estimates as response variable against all environmental and data characteristic variables

Generalized Linear model output for Chao1 Richness Estimates

Characteristic	Both substrates			Litter			Soil		
	Beta	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**	**Beta**	**95% CI**[1]	**p-value**
Substrate									
Litter	—	—							
Soil	-37	-109, 34	0.3						
Vegetation type									
Alpine heath	—	—		—	—		—	—	
Birch	35	-66, 137	0.5	91	-50, 232	0.2	-11	-161, 139	0.9
Broad leaf	-3.9	-186, 179	>0.9	78	-171, 327	0.5	-49	-339, 242	0.7
Mix	56	-118, 230	0.5	223	-30, 477	0.09	-65	-307, 177	0.6
Pine	26	-113, 166	0.7	126	-66, 317	0.2	-78	-285, 129	0.5
Polar	-496	-952, -41	**0.034**	14	-646, 673	>0.9	-726	-1,389, -63	**0.036**
Ridge	32	-62, 126	0.5	29	-100, 158	0.7	77	-70, 223	0.3
Snowbed	33	-73, 139	0.5	-30	-186, 125	0.7	118	-50, 285	0.2
Spruce	61	-80, 201	0.4	136	-61, 333	0.2	-7.8	-210, 194	>0.9
N	-199	-596, 198	0.3	70	-1,050, 1,190	>0.9	-233	-925, 458	0.5
C	167	-154, 488	0.3	41	-846, 928	>0.9	20	-571, 611	>0.9
C:N	-130	-486, 225	0.5	-36	-1,049, 978	>0.9	46	-622, 713	0.9
P	-45	-237, 147	0.6	105	-223, 432	0.5	-114	-379, 150	0.4
pH	-261	-566, 45	0.1	-302	-851, 248	0.3	-561	-1,026, -97	**0.021**
Latitude	-6	-408, 396	>0.9	196	-380, 772	0.5	-65	-638, 509	0.8
Longitude	-256	-587, 74	0.13	-251	-698, 196	0.3	-230	-715, 255	0.4
Metres above sea level	-216	-535, 102	0.2	150	-319, 618	0.5	-487	-947, -27	**0.043**
Annual mean temperature	-1,578	-4,493, 1,338	0.3	-1,576	-5,803, 2,650	0.5	-901	-5,131, 3,328	0.7
Mean diurnal temp. range	-1,124	-2,354, 106	0.075	-192	-1,927, 1,543	0.8	-1,545	-3,323, 233	0.094
Isothermality	484	-158, 1,126	0.14	69	-830, 968	0.9	719	-228, 1,666	0.14
Temperature seasonality	673	-294, 1,640	0.2	1,120	-258, 2,498	0.12	34	-1,363, 1,431	>0.9
Mean temp. of Wettest Quarter	169	-32, 370	0.1	17	-269, 304	>0.9	149	-156, 454	0.3
Mean temp. of Driest Quarter	138	-188, 465	0.4	-170	-643, 302	0.5	308	-185, 802	0.2
Mean temp. of Warmest Quarter	686	-1,209, 2,580	0.5	59	-2,760, 2,877	>0.9	938	-1,910, 3,787	0.5
Mean temp. of Coldest Quarter	567	-1,430, 2,563	0.6	2,211	-532, 4,954	0.12	-1,016	-3,930, 1,898	0.5
Annual precipitation	-216	-649, 218	0.3	91	-511, 694	0.8	-474	-1,150, 202	0.2
Precipitation seasonality	-152	-400, 97	0.2	-57	-418, 304	0.8	-245	-603, 113	0.2
Plant richness	41	-133, 214	0.6	156	-85, 398	0.2	59	-216, 334	0.7
Sample read abundance	0.01	0.00, 0.01	**0.007**	0.01	0.00, 0.01	**0.004**	0	0.00, 0.01	0.5
% unique OTUs in sample	16	10, 23	**<0.001**	14	1.9, 26	**0.027**	18	8.3, 27	**<0.001**
No. Obs.	168			84			84		
Null deviance	4,081,323			1,764,665			2,276,017		
Null df	167			83			83		
Deviance	2,593,928			941,145			1,038,469		
Residual df	137			54			54		

[1] CI = Confidence Interval

p < 0.05

5. Beta diversity statistical analysis summaries (PERMANOVAs)

Table S10: Summary table for PERMANOVA outputs on Bray-Curtis distance matrices

	PERMANOVA output																				
	Both substrates							Litter							Soil						
Character		Df	SumOfSqs	R^2	F	Pr(>F)		Df	SumOfSqs	R^2	F	Pr(>F)			Df	SumOfSqs	R^2	F	Pr(>F)		
Vegetation type	forest_type	8	10.128	0.15311	3.6383	1e-04	***	8	7.631	0.23986	2.9977	1e-04	***		8	6.3463	0.21284	2.5686	1e-04	***	
	Residual	167	62.402	0.94337				76	24.184	0.76014					76	23.4715	0.78716				
Carbon content	C	1	2.981	0.04506	7.928	1e-04	***	1	0.944	0.02966	2.5372	1e-04	***		1	1.5282	0.05125	4.4837	1e-04	***	
	Residual	168	63.167	0.95494				83	30.872	0.97034					83	28.2896	0.94875				
Nitrogen content	N	1	2.650	0.04006	7.01	1e-04	***	1	0.779	0.0245	2.0844	3e-04	***		1	0.7791	0.02613	2.2268	6e-04	***	
	Residual	168	63.498	0.95994				83	31.036	0.9755					83	29.0387	0.97387				
Carbon:Nitrogen ratio	C_N	1	2.211	0.03343	5.8106	1e-04	***	1	1.075	0.0338	2.9037	1e-04	***		1	1.9716	0.06612	5.8768	1e-04	***	
	Residual	168	63.937	0.96657				83	30.74	0.9662					83	27.8461	0.93388				
Phosphorus content	P	1	1.664	0.02515	4.3348	1e-04	***	1	0.823	0.02587	2.2043	1e-04	***		1	0.6024	0.0202	1.7114	0.0064	**	
	Residual	168	64.484	0.97485				83	30.993	0.97413					83	29.2154	0.9798				
pH	pH	1	1.688	0.02587	4.4083	1e-04	***	1	1.0422	0.03315	2.8114	1e-04	***		1	1.3788	0.04706	4.0491	1e-04	***	
	Residual	166	63.547	0.97413				82	30.3961	0.96685					82	27.9221	0.95294				
Plant richness	ılant_richnes:	1	0.742	0.01122	1.9058	0.0012	**	1	0.438	0.01378	1.1596	0.1772			1	0.6461	0.02167	1.8384	0.0038	**	
	Residual	168	65.406	0.98878				83	31.377	0.98622					83	29.1716	0.97833				
Metres above sea level	MAS	1	1.903	0.02876	4.9752	1e-04	***	1	1.487	0.04673	4.069	1e-04	***		1	1.0952	0.03673	3.1649	1e-04	***	
	Residual	168	64.245	0.97124				83	30.329	0.95327					83	28.7226	0.96327				
Longitude	long_zs	1	1.100	0.01662	2.8402	1e-04	***	1	0.660	0.02076	1.7593	0.0033	**		1	0.7982	0.02677	2.2829	4e-04	***	
	Residual	168	65.048	0.98338				83	31.155	0.97924					83	29.0196	0.97323				
Latitude	lat_zs	1	1.561	0.02359	4.0594	1e-04	***	1	1.155	0.03629	3.1255	1e-04	***		1	1.0511	0.03525	3.0328	1e-04	***	
	Residual	168	64.587	0.97641				83	30.661	0.96371					83	28.7667	0.96475				
Annual mean temp.	bio_1	1	2.200	0.03325	5.779	1e-04	***	1	1.738	0.05461	4.7946	1e-04	***		1	1.3242	0.04441	3.8574	1e-04	***	
	Residual	168	63.948	0.96675				83	30.078	0.94539					83	28.4935	0.95559				
Mean diurnal temp. range	bio_2	1	1.187	0.01795	3.0709	1e-04	***	1	0.829	0.02606	2.2206	2e-04	***		1	0.7458	0.02501	2.1291	0.0011	**	
	Residual	168	64.961	0.98205				83	30.987	0.97394					83	29.0720	0.97499				
Isothermality	bio_3	1	1.201	0.01815	3.1061	1e-04	***	1	0.835	0.02624	2.2368	1e-04	***		1	0.8827	0.0296	2.5321	1e-04	***	
	Residual	168	64.947	0.98185				83	30.981	0.97376					83	28.9350	0.9704				
Temp. seasonality	bio_4	1	0.918	0.01388	2.3649	2e-04	***	1	0.717	0.02252	1.9124	0.0011	**		1	0.5349	0.01794	1.516	0.0208	*	
	Residual	168	65.23	0.98612				83	31.099	0.97748					83	29.2829	0.98206				
Mean temp. of wettest quarter	bio_8	1	1.418	0.02143	3.679	1e-04	***	1	0.938	0.02949	2.5216	1e-04	***		1	0.9564	0.03208	2.7505	1e-04	***	
	Residual	168	64.731	0.97857				83	30.878	0.97051					83	28.8614	0.96792				
Mean temp. of driest quarter	bio_9	1	1.124	0.01699	2.9041	1e-04	***	1	1.019	0.03203	2.7461	1e-04	***		1	0.6389	0.02143	1.8173	0.0026	**	
	Residual	168	65.024	0.98301				83	30.797	0.96797					83	29.1789	0.97857				
Mean temp. of warmest quarter	bio_10	1	2.393	0.03618	6.307	1e-04	***	1	1.717	0.05397	4.7352	1e-04	***		1	1.5703	0.05266	4.6139	1e-04	***	
	Residual	168	63.755	0.96382				83	30.098	0.94603					83	28.2475	0.94734				
Mean temp. of coldest quarter	bio_11	1	1.724	0.02607	4.4962	1e-04	***	1	1.449	0.04555	3.9609	1e-04	***		1	0.9702	0.03254	2.7914	1e-04	***	
	Residual	168	64.424	0.97393				83	30.366	0.95445					83	28.8476	0.96746				
Annual precipitation	bio_12	1	1.048	0.01585	2.7057	1e-04	***	1	0.798	0.02509	2.136	5e-04	***		1	0.6581	0.02207	1.8731	0.0035	**	
	Residual	168	65.1	0.98415				83	31.017	0.97491					83	29.1597	0.97793				
Precipitation seasonality	bio_15	1	0.847	0.0128	2.1787	7e-04	***	1	0.729	0.02291	1.9463	6e-04	***		1	0.4966	0.01665	1.4057	0.0486	*	
	Residual	168	65.301	0.9872				83	31.087	0.97709					83	29.3212	0.98335				
Sample site	SITE	89	41.185	0.62261	1.483	1e-04	***														
	Residual	80	24.963	0.37739																	
Substrate type	TYPE	1	4.637	0.0701	12.665	1e-04	***														
	Residual	168	61.511	0.9299																	
ɔdes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1																					