Received: 27 September 2021 | Revised: 1 February 2022 | Accepted: 16 February 2022

DOI: 10.1111/geb.13487

Global Ecology and Biogeography | A Journal of Macroecology

WILEY

RESEARCH ARTICLE

Global soil microbiomes: A new frontline of biome-ecology research

Martti Vasar[1] | John Davison[1] | Siim-Kaarel Sepp[1] | Ladislav Mucina[2,3] | Jane Oja[1] | Saleh Al-Quraishy[4] | Sten Anslan[1] | Mohammad Bahram[5] | C. Guillermo Bueno[1] | Juan José Cantero[6,7] | Guillaume Decocq[8] | Lauchlan Fraser[9] | Inga Hiiesalu[1] | Wael N. Hozzein[4,10] | Kadri Koorem[1] | Yiming Meng[1] | Mari Moora[1] | Vladimir Onipchenko[11] | Maarja Öpik[1] | Meelis Pärtel[1] | Tanel Vahter[1] | Leho Tedersoo[1] | Martin Zobel[4,12]

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

[2]Harry Butler Institute, Murdoch University, Murdoch, Perth, Western Australia, Australia

[3]Department of Geography & Environmental Studies, Stellenbosch University, Stellenbosch, South Africa

[4]Zoology Department, College of Science, King Saud University, Riyadh, Saudi Arabia

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

[6]Universidad Nacional de Córdoba, Instituto Multidisciplinario de Biología Vegetal, CONICET, Córdoba, Argentina

[7]Departamento de Biología Agrícola, Facultad de Agronomía y Veterinaria, Universidad Nacional de Río Cuarto, Córdoba, Argentina

[8]Ecologie et Dynamique des Systèmes Anthropisés, Jules Verne University of Picardie, Amiens, France

[9]Department of Natural Resource Sciences, Thompson Rivers University, Kamloops, British Columbia, Canada

[10]Botany and Microbiology Department, Faculty of Science, Beni-Suef University, Beni-Suef, Egypt

[11]Department of Ecology and Plant Geography, Faculty of Biology, Moscow Lomonosov State University, Moscow, Russia

[12]Department of Botany, University of Tartu, Tartu, Estonia

Correspondence
John Davison, Institute of Ecology and Earth Sciences, University of Tartu, Estonia.
Email: john.davison@ut.ee

Funding information
Vetenskapsrådet, Grant/Award Number: 2017-05019; Eesti Teadusagentuur, Grant/Award Number: MOBTP 105, PRG 1065, PRG 1170 and PRG 609; Tartu Ülikool, Grant/Award Number: PLTOM20903; European Regional Development Fund; Russian Ministry of Science and Higher Education, Grant/Award Number: 075-15-2021-1396

Handling Editor: Janne Soininen

Abstract

Aim: Organisms on our planet form spatially congruent and functionally distinct communities, which at large geographical scales are called "biomes". Understanding their pattern and function is vital for sustainable use and protection of biodiversity. Current global terrestrial biome classifications are based primarily on climate characteristics and functional aspects of plant community assembly. These and other existing biome schemes do not take account of soil organisms, including highly diverse and functionally important microbial groups. We aimed to define large-scale structure in the diversity of soil microbes (soil microbiomes), pinpoint the environmental drivers shaping it and identify resemblance and mismatch with existing terrestrial biome schemes.

Location: Global.

Time period: Current.

Major taxa studied: Soil eukaryotes and prokaryotes.

Methods: We collected soil samples from natural environments world-wide, incorporating most known terrestrial biomes. We used high-throughput sequencing to

Martti Vasar, John Davison, Siim-Kaarel Sepp, Ladislav Mucina, Leho Tedersoo and Martin Zobel contributed equally to this work.

characterize soil biotic communities and *k*-means clustering to define soil microbiomes describing the diversity of microbial eukaryotic and prokaryotic groups. We used climatic data and soil variables measured in the field to identify the environmental variables shaping soil microbiome structure.

Results: We recorded strong correlations among fungal, bacterial, archaeal, plant and animal communities, defined a system of global soil microbiomes (producing seven biome types for microbial eukaryotes and six biome types for prokaryotes) and showed that these are typically structured by pH alongside temperature. None of the soil microbiomes are directly paralleled by any current terrestrial biome scheme, with mismatch most substantial for prokaryotes and for microbial eukaryotes in cold climates; nor do they consistently distinguish grassland and forest ecosystems.

Main conclusions: Existing terrestrial biome classifications represent a limited surrogate for the large-scale diversity patterns of microbial soil organisms. We show that empirically defined soil microbiomes are attainable using metabarcoding and statistical clustering approaches and suggest that they can have wide application in theoretical and applied biodiversity research.

KEYWORDS
biodiversity, biogeography, metabarcoding, pH, soil biota

1 | INTRODUCTION

Biomes are large-scale biotic communities derived from criteria relevant to their structure and functioning. Current approaches to classification of aboveground terrestrial biomes use macroclimatic drivers, on the assumption that these underpin the major physiognomic appearance (well-defined vegetation formations and associated animal communities) and structure (spatial arrangement of biotic elements) of biomes and, as a consequence, their functioning, such as responses to environmental stresses and disturbances and to water and nutrient dynamics (Mucina, 2019; Olson et al., 2001). As such, biomes provide a broad description of how life on Earth is structured. Biomes are popular in teaching of ecology and environmental sciences, provide large-scale context to ecological and biogeographical studies (e.g., Bastida et al., 2020) and are tools for large-scale conservation planning (Brooks et al., 2006), generalizing human impacts upon landscapes (Jacobson et al., 2019) and forecasting the impact of global change (Boit et al., 2016). The functional aspect of biomes provides a mechanistic basis for explanatory and predictive modelling and for development of the link between biogeography and ecosystem functionality, including provision of ecosystem services and land use (Delgado-Baquerizo, Reich, Trivedi, et al., 2020; Moncrieff et al., 2016).

It is widely acknowledged that soil microbes represent a substantial fraction of global biodiversity and are fundamental to vegetation and ecosystem structure and functioning (e.g., Guerra et al., 2020; Xu et al., 2020). Yet, microbial communities have been almost entirely neglected in attempts to define or validate terrestrial biomes. Smith et al. (2018), who validated "ecoregions" (geographically distinct subunits within biomes) using plant, fungal and animal database records, found that the distributions of fungi did not adhere to ecoregion boundaries. There is also evidence that certain microbial groups are relatively more sensitive to ecological gradients than to biogeographical constraints, in comparison with macro-organisms (Tipton et al., 2019), and respond to different environmental drivers (including strong effects of certain edaphic properties) from those defining vegetation patterns (Shen et al., 2014). It therefore seems likely that consideration of microbial diversity will fundamentally change the way we view the large-scale structure of life on Earth.

Although compiling the distributions of regionally restricted macro-organisms poses an almost insurmountable logistic challenge, metabarcoding techniques coupled with expert definition of functional types provides a more viable approach for characterizing ecologically meaningful units within soil microbial diversity. Consequently, there appears to be a strong argument for deriving soil microbial biomes (hereafter, soil microbiomes) directly from microbial datasets, rather than using terrestrial biomes as a surrogate or stratification tool.

Here, we present a dataset comprising soil samples from all continents except Antarctica. Predominantly microbial soil biota were described by sequencing eukaryotic and prokaryotic DNA, and major guilds and taxonomic divisions were distinguished among fungi, prokaryotes, plants and animals. We generated global soil microbiome classifications based on these empirical data and aimed to determine the extent to which current terrestrial biomes reflect variation in different soil organism groups; and how soil microbiome structure, and mismatch with terrestrial biomes, are shaped by edaphic properties alongside climate.

2 | MATERIALS AND METHODS

2.1 | Sample collection

Three hundred and forty-five soil samples from previously described sampling locations world-wide (Davison et al., 2021) were used in this study (Supporting Information Dataset S1). Sampling locations generally experienced little disturbance from human activities, and sampling followed the approach described by Tedersoo et al. (2014): *c*. 300 g of topsoil was collected from ≤40 points within a sampling area of *c*. 50 m × 50 m and pooled. Soil samples were dried within 24 h using silica gel or at room temperature, then carefully homogenized. Subsamples (2 g) of soil were extracted for molecular analysis; the remainder was stored for soil chemical analysis.

2.2 | Environmental variables

Soil samples were sieved using a 2 mm mesh before analysis of soil chemical properties: pH, total N, organic C, P, Mg and K. Soil pH was measured in 1 M KCl solution following ISO 10390:2005 (Seven Easy pH meter with InLab Expert Pro electrode; Mettler Toledo, Malaysia). Total N content was determined using the Kjeldahl method (DK-20 digestion block and UDK-126 distillation unit; Velp Scientifica, Italy). Organic carbon content was determined using Tjurin's method (oxidation provided by boiling in $H_2SO_4/K_2Cr_2O_7$ solution) and subsequent spectrophotometry (Specol-11; Carl Zeiss, Germany). The Mehlich III extraction method was used to determine P, K, Mg and Ca content (MP-4200 microwave plasma atomic emission spectrometer; Agilent, USA). Soil chemical analyses were performed at the Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Tartu, Estonia. Estimates of mean annual temperature (MAT), mean annual precipitation (MAP) and precipitation seasonality (SeaPrec) at sample locations were taken from the CHELSA database (Karger et al., 2017). A small number of missing values (two pH, four P, 10 K and 10 Ca) were assigned the mean value across the dataset in order to retain all data points in analyses (Supporting Information Dataset S1). The historical stability of terrestrial biomes at sampling locations was estimated by comparing Olson's current biome classification (Olson et al., 2001) with an analogous classification for the Last Glacial Maximum (*c*. 21 kyr BP; Ray & Adams, 2001). Where the biome remained the same at both time points, samples were classified as stable; where the biome classification was different, samples were classified as unstable.

2.3 | Molecular methods

DNA was extracted from 2 g of dried soil using the PowerMax Soil DNA Isolation Kit. The internal transcribed spacer (ITS) region was used for identification of soil microbial eukaryotes; the 16S rRNA gene was used for identification of soil prokaryotes. Certain macro-organism groups were also identified: plants were also identified using ITS, whereas soil animals were identified using the 18S rRNA gene. Given that the sampling was designed to target microorganisms in topsoil, the plant and animal data were expected to be less complete and provide only partial information.

2.3.1 | Eukaryotes

The ITS region was amplified using primers ITS9mun (5′-TGTACACACCGCCCGTCG-3′) and ITS4ngsUni (5′-CCTSCSCTTANTDATATGC-3′) for identification of eukaryotes in general, and the 18S rRNA gene was amplified with primer pair Euk575F (5′-ASCYGYGGTAAYWCCAGC-3′) and Euk895R (5′-TCHNHGNATTTCACCNCT-3′) for identification of non-fungal eukaryotes, specifically soil animals. The primer pair ITS9mun and ITS4ngsUni has been proposed for analysis of eukaryotes with approximately species-level resolution (Tedersoo & Anslan, 2019; Tedersoo & Lindahl, 2016). These primers amplify 170 bp of the 18S rRNA gene and the full-length ITS region, which serves as an official barcode for fungi (Nilsson et al., 2018; Schoch et al., 2012) and performs well on many other groups, including plants (Coleman, 2009; Pawlowski et al., 2012). All primers were equipped with unique 12-base Golay barcodes for multiplexing.

For amplification, the polymerase chain reaction (PCR) was performed in two replicates for each sample, comprising 5 μl of 5× HOT FIREPol Blend Master Mix (Solis Biodyne, Tartu, Estonia), 0.5 μl of each forward and reverse primer (20 mM), 1 μl of DNA extract and 18 μl ddH_2O. Thermal cycling included an initial hot-start denaturation at 95°C for 15 min; 30 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 57°C and elongation for 1 min at 72°C; and final elongation at 72°C for 10 min and storage at 4°C. The duplicate PCR products were pooled and checked for the presence of amplification product on 1% agarose gel. In the case of no bands or weakly visible bands, samples were reamplified with higher numbers of PCR cycles (maximum of 35 cycles). Negative and positive controls were used throughout the process.

The PCR products were purified using a FavorPrepTM GEL/PCR Purification Kit (Favorgen Biotech Corporation). The ITS eukaryote libraries were sequenced on a PacBio Sequel instrument using SMRT cell 1 M, v.2 LR; Sequel Polymerase v.2.1 and Sequencing chemistry v.2.1. Loading was performed by diffusion; one SMRT cell was used for sequencing, with a movie time of 600 min and a pre-extension time of 45 min. The use of PacBio technology to sequence a large fragment of this marker region represents an important advance compared with earlier global metabarcoding studies, and it is notable that PacBio is considerably less prone to bias regarding ITS length variation than other technologies (Castaño et al., 2020). The 18S rRNA gene libraries of non-fungal eukaryotes were ligated with Illumina adaptors using the TruSeq DNA PCR-free library prep kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina MiSeq platform, using a 2 × 250 bp paired-read sequencing approach, at Asper Biogene (Tartu, Estonia).

2.3.2 | Prokaryotes

Prokaryotic primers 515F (5′-GTGYCAGCMGCCGCGGTAA-3′) and 926R (5′-GGCCGYCAATTYMTTTRAGTTT-3′) were used to amplify the 16S rRNA variable V4 region (Caporaso et al., 2011; Parada et al., 2016). Both primers were equipped with unique 12-base Golay barcodes for multiplexing.

The PCR was performed in two replicates per sample in a 25 μl reaction volume containing 5 μl of 5× HOT FIREPol Blend MasterMix (Solis Biodyne), 0.5 μl of both forward and reverse primer (20 mM) and 1 μl of the DNA extract (the remaining volume was filled with nuclease-free water). The PCR amplification program included the following steps: 95°C for 15 min, followed by 25 cycles of 95°C for 30 s, 55°C 30 s and 72°C for 1 min, with a final extension step at 72°C for 10 min. The duplicate PCR products were pooled and visualized on 1% agarose gel. Initially, 25 cycles were used for all the samples, and in cases where the gel band was weak or there was no PCR product a higher number of PCR cycles was used (maximum of 30 cycles). Negative and positive controls were included throughout the process.

The PCR products were purified using a FavorPrepTM GEL/PCR Purification Kit (Favorgen Biotech Corporation). The libraries were ligated with Illumina adaptors using the TruSeq DNA PCR-free library prep kit (Illumina). Libraries were sequenced on the Illumina MiSeq platform, using a 2 × 250 bp paired-read sequencing approach, at Asper Biogene (Tartu, Estonia).

Raw reads from this targeted locus study have been deposited in the NCBI SRA (BioProject PRJNA659159).

2.4 | Bioinformatics

2.4.1 | Eukaryotes

ITS PacBio CCS reads (minPasses = 3, MinAccuracy = 0.9) were generated using SMRT Link v.6.0.0.47841. Subsequent quality filtering was performed using PipeCraft v.1.0 (Anslan et al., 2017) as described by Tedersoo and Anslan (2019), yielding 3,544,553 cleaned reads. Flanking regions (SSU and LSU) were extracted and removed using ITSx (v.1.1.3; Bengtsson-Palme et al., 2013), leaving 2,425,323 reads. UNITE v.8.2 (all eukaryotes; Kõljalg et al., 2013) served as a reference database for chimera filtering and identification. Chimera checking retained 2,189,815 potential chimera-free sequences. Using VSEARCH (v.2.14.1; Rognes et al., 2016), sequences were clustered into 98,483 operational taxonomic units (OTUs; singletons removed) at 98% sequence similarity using the abundance option to allow separation of closely related species while keeping lower-quality sequences and rare variants adhered to OTU centroids. Global singletons were removed.

For taxonomic assignment, we evaluated the 10 best hits and conservatively kept OTUs with conflicting best matches unidentified at the level of that taxonomic rank. BLAST (v.2.11.0+; Camacho et al., 2009) hits were evaluated using an alignment criterion of 80% of the shorter read (query or reference) to remove putative chimeric reads that passed the chimera checking. BLAST against UNITE identified 57,679 OTUs (1,796,814 reads). Sequences were assigned to taxonomic levels as follows: orders, families and genera were assigned at >80, >85 and >90% sequence similarity, respectively. Fungal taxa [taxonomy follows Tedersoo et al. (2018) as updated by Wijayawardene et al. (2020)] were used at the level of phyla (early diverging lineages), classes (Chytridiomycota, most Mucoromycota) and orders (Dikarya, moulds) to balance between taxonomic resolution and coverage.

To assign functions [arbuscular mycorrhizal (AM), ectomycorrhizal (EcM), multitroph, pathogen and saprotroph] to fungal OTUs, we took two parallel approaches. First, we used the newly built FungalTraits database (Põlme et al., 2020) to assign OTUs to guilds and EcM fungi further to lineages and exploration types. For genera with multiple lifestyles, we used the assignments based on annotations at the level of sequences and species hypotheses (SHs) as given in UNITE. A fungal function was assigned to 44,258 OTUs using FungalTraits (AM: 4,645 OTUs, 35,615 reads; EcM: 7,754 OTUs, 457,701 reads; multitroph: 9,067 OTUs, 299,120 reads; pathogen: 3,323 OTUs ,81,127 reads; saprotroph: 19,469 OTUs, 639,066 reads). Other microbial eukaryotic organisms (2,711 OTUs, 149,061 reads) and plants (1,495 OTUs, 58,066 reads) were also retained; the remaining 9,215 OTUs (77,058 reads) that achieved only low identity or alignment (<80%) BLAST hits against the groups of interest, or were not matched against the FungalTraits database (for fungal OTUs), were discarded. Plant OTUs were distinguished into Tracheophyta, algae and Bryophyta. Sequence counts and functional annotations for eukaryotic ITS OTUs are given in the Supporting Information (Dataset S2).

The 18S rRNA gene animal data were analysed using the gDAT pipeline (Vasar et al., 2021). Demultiplexed paired-end reads were analysed in the following way: barcode and primer sequences were matched, allowing one mismatch for both pairs. Only pairs where both reads had an average quality score of >30 were retained (after removal of barcode and primer sequences). Quality filtered paired-end reads were combined using FLASH (v.1.2.10; Magoč & Salzberg, 2011) with the default parameters (10–300 bp overlap with ≥75% identity). Orphan reads (paired-end reads that did not meet the conditions for combination) were removed from the analyses, leaving 18,492,667 cleaned sequences. The VSEARCH chimera filtering algorithm was used to remove putative chimeric reads in *de novo* mode, yielding 18,104,321 chimera-free sequences. Reads were clustered with VSEARCH at 98% identity into 82,736 OTUs (excluding singletons). Representative sequences (OTU centroids) for each non-singleton OTU were classified taxonomically using a BLAST search followed by selection of the best hit against the well-curated Protist Ribosomal Reference database (v.4.12.0; Guillou et al., 2012), resulting in 49,455 hits. Animals were filtered based on three phylum groups (Annelida, Nematoda and Arthropoda), resulting in 4,185 OTUs and 1,258,416 reads. Samples with >1,000 reads were retained for analysis. Sequence counts and functional annotations for animal 18S OTUs are given in the Supporting Information (Dataset S3).

2.4.2 | Prokaryotes

Bioinformatic treatment of 16S rRNA gene prokaryotic data followed the protocol used for animal data, as described above, aside from the following details. Quality filtering and cleaning left 14,593,006 cleaned sequences; chimera filtering yielded 13,960,691 chimera-free sequences that were clustered at 99% identity (following the recommendation by Edgar, 2018) into 165,330 OTUs (excluding singletons). Representative sequences (OTU centroids) for each non-singleton OTU were classified taxonomically using a BLAST search against the SILVA database (v.132; Pruesse et al., 2007), taking the lowest common ancestor, resulting in 141,959 hits. Prokaryotic OTUs (141,828 OTUs, 4,935,514 reads) were distinguished into bacteria (140,343 OTUs, 4,859,933 reads) and archaea (1,485 OTUs 75,581 reads). The remaining 23,502 OTUs (9,025,117 reads), for which taxonomic resolution was low or missing (lower than family), were omitted. Functional profiles for prokaryotic taxa (C-cycle, chemotrophs, chemoheterotrophs, N-cycle and parasites) were mapped to 44,857 OTUs using FAPROTAX (v.1.2.4; Louca et al., 2016). Sequence counts and functional annotations for prokaryotic 16S OTUs are given in the Supporting Information (Dataset S4).

2.5 | Statistical methods

2.5.1 | Data transformation

Preliminary rarefying of samples with high read counts indicated that full sample composition could be estimated accurately by subsampling ≥1,000 reads. Therefore, this sample size was used as a minimum threshold in each data matrix, and samples with <1,000 reads were excluded from the rest of the analysis. To normalize the remaining sequence count data, we implemented the variance stabilizing transformation (using the R package DESeq2 v.1.28.1; Love et al., 2014), as suggested by McMurdie and Holmes (2014). The method uses fitted dispersion–mean relationships to normalize data with respect to sample size (sequencing depth of individual samples) and variance.

2.5.2 | Large-scale pattern in soil biodiversity

We used unsupervised *k*-means clustering to identify global patterns in the composition of a number of soil organism groups: prokaryotes, in addition to bacteria and archaea separately; microbial eukaryotes, in addition to fungi and component fungal guilds (AM, EcM, multitrophic, pathogenic and saprotrophic) separately; and plant and animal macro-organism groups.

For each group, principal coordinates analysis (PCoA) vectors derived from a Bray–Curtis distance matrix were included in the *k*-means clustering algorithm. Clustering with values of k = 2–16, with 100 random starting configurations per *k*, was used. Silhouette plots (Supporting Information Figure S1) were used to ascertain the best-fitting cluster configurations among the range of *k*-values. Bootstrapping was used to assess the stability of clustering solutions (Hennig, 2007), with Jaccard similarity of empirical and bootstrapped clustering configurations calculated (Supporting Information Figure S1). For microbial eukaryotic (ITS) communities, the best-supported values were k = 2, 3, 4 and 7, and for prokaryotes, k = 2, 3, 4 and 6. We focused on the higher values of *k* for comparison with existing biome classifications.

Cluster maps were interpolated using weighted categorical *k*-nearest neighbour (KNN) classification [kknn() in the R package kknn; Schliep & Hechenbichler, 2016], using the soil sample cluster identities as the training set and a 0.5° × 0.5° map grid as the test set. The weights were based on great-circle distances of the *k* nearest training set points from the respective test set grid point. Grid cell cluster identity represents the class with the highest probability in the KNN prediction at that particular point. The *k*-value (k = 18) for KNN was set as the rounded square root of the number of samples, based on the suggestion of Duda et al. (2012). Greenland and the Sahara region were excluded from the interpolation owing to insufficient sampling and contrastingly different abiotic conditions. The composition of microbiome classes was calculated in relationship to the fungal (AM, EcM, multitrophic, pathogenic and saprotrophic) and prokaryotic guilds (C-cycle, chemotrophs, chemoheterotrophs, N-cycle and parasites) and the plant (Tracheophyta, algae and Bryophyta) and animal (Annelida, Arthropoda and Nematoda) taxonomic groupings.

2.5.3 | Drivers of global soil communities

Abiotic drivers of eukaryotic and prokaryotic soil communities were identified using distance-based redundancy analysis (dbRDA; "vegan" package in R; Oksanen et al., 2020) and generalized dissimilarity modelling (gdm; Guerin et al., 2021). dbRDA can incorporate categorical predictor variables, whereas gdm allows nonlinear effects to be identified and a spatial distance matrix to be incorporated directly. In the dbRDA analyses, the variation in sample distance matrices (Bray–Curtis distance following variance stabilizing transformation) was modelled against measured abiotic variables. A set of independent predictor variables for inclusion in models was selected based on pairwise correlations and variance inflation factors (VIFs) in full models, retaining the variables that exhibited the strongest explanatory power. Thus, we removed N (correlated with organic carbon, r = 0.83) and Mg (correlated with Ca, r = 0.5) from final analyses (all variable VIF values for final models ≤2.64). The significance of effects was measured using permutation (n = 999). In gdm models, all continuous abiotic variables were included along with a great earth circle distance matrix representing spatial distances between sampling locations. The importance and significance of variables in gdm models were measured by calculating the change in deviance between full models of intact data and models calculated after permutation of the

variable of interest. Nonlinear effects were visualized by plotting variable I-splines.

Correlation between organism guilds was measured using Procrustes correlation (the Procrustes function from vegan). For each guild, a Bray–Curtis distance matrix was decomposed into a series of vectors using PCoA, and the two matrices of PCoA vectors were included in the Procrustes analysis. In the analysis, only vectors with positive eigenvalues were used, and where the number of positive vectors differed between matrices, the minimum number of vectors was included for both sets. The significance of correlation was assessed using permutation (n = 999). Variation in Procrustes residuals, which indicates the local strength of correlation, was analysed in relationship to the measured abiotic variables using linear models.

3 | RESULTS

3.1 | Soil microbiomes

For microbial eukaryotes, k-means clustering (k = 7) revealed two series of community clusters, arranged along the temperature gradient (Figure 1a): clusters "cold acid" → "cool acid" → "temperate acid wet" → "tropical acid wet" reflecting increasing temperature in acid soil conditions; and clusters "cool basic" → "temperate basic dry" → "tropical basic dry" reflecting increasing temperature in basic soil conditions (in warmer conditions, the clusters also diverged with respect to precipitation conditions). The broader microbial eukaryotic clusters (lower values of k) primarily represented climatic zones with some modification owing to the impact of soil pH (Supporting Information Figure S2a–c). Prokaryotic k = 6 clustering revealed two series of clusters arranged along the pH gradient (Figure 1b): clusters "cool acid" → "cool intermediate" → "cool basic" reflecting increasing pH in cool climates; and clusters "warm acid" → "warm intermediate" → "warm basic" reflecting increasing pH in warm climates. The broader prokaryotic clusters (lower values of k) primarily represented edaphic types, with modifications owing to the impact of temperature (Supporting Information Figure S2d–f). Fungal guild clusters largely matched the overall eukaryotic classification; bacterial and archaeal clusters largely matched the overall prokaryotic classification; and both plant and animal clusters represented complex combinations of climatic and edaphic conditions, with some regional differentiation (Supporting Information Figure S3).

Guild structure varied significantly between clusters (Figure 2; Supporting Information Figure S4), with microbial eukaryotic clusters "cool basic" and "temperate acid wet" characterized by the

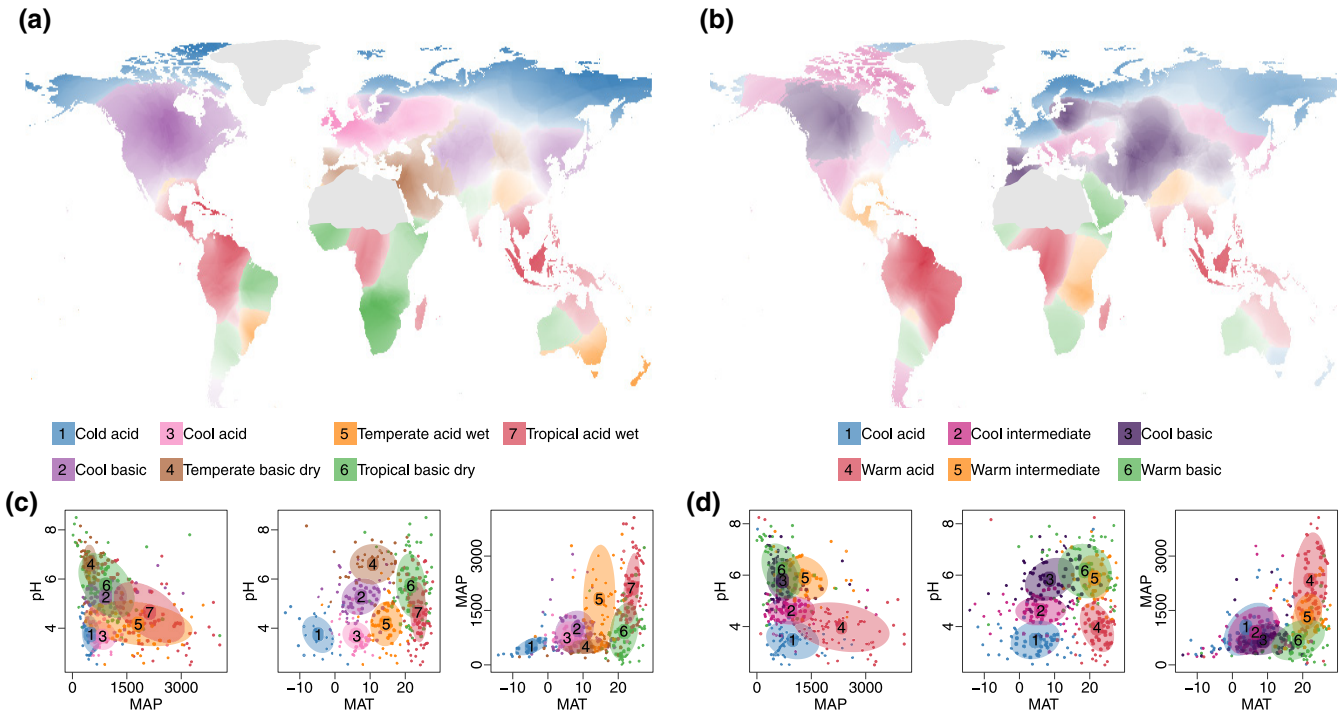

FIGURE 1 Clusters of global soil biodiversity. (a,b) Global maps showing the distributions of clusters of (a) microbial eukaryotic and (b) prokaryotic soil communities. Clusters were defined using k-means clustering [k = 7 for microbial eukaryotes (ITS); k = 6 for prokaryotes (16S)], and cluster distribution was interpolated using k-nearest neighbour (k = 18) spatial interpolation. Colour intensity denotes the confidence of the assignment for the most credible cluster identity. Greenland and the Sahara region were excluded from the maps because of insufficient sampling and contrastingly different abiotic conditions. (c,d) Positions of clusters along axes of pH, MAT and MAP for (c) microbial eukaryotes and (d) prokaryotes, respectively. Ellipses around cluster centroids are standard deviational ellipses. Note that plants and soil animals (18S) are not included in the soil microbial eukaryotic calculations. MAP = mean annual precipitation; MAT = mean annual temperature

highest and the cluster "tropical acid dry" by the lowest proportional abundance of EcM fungi, and cluster "temperate basic dry" by the highest abundance of AM fungi. The northern-most plant community cluster contained the highest share of bryophytes (cluster 1; Supporting Information Figures S4 and S5). Prokaryotic communities were dominated by chemoheterotrophs, but there was a higher share of C-cycle-related prokaryotes in the cluster "cool acid" and of N-cycle related prokaryotes in the cluster "warm intermediate" (Supporting Information Figures S2 and S4). Among animal communities, there was a higher share of Nematoda in a widespread Northern Hemisphere cluster (cluster 1) and of Arthropoda in a cluster with a broadly subtropical distribution (cluster 4; Supporting Information Figures S4 and S5). Notably, consistent differences in vegetation structure (e.g., forests vs. grasslands), which are defining criteria for existing biome classifications, were not apparent in microbial eukaryotic or prokaryotic clusters (Figure 3; Supporting Information Figure S4).

3.2 | Drivers of soil microbiome structure

We recorded strong correlations in community composition between microbial eukaryotes, prokaryotes, plants and animals, in addition to their component guilds (pairwise Procrustes correlations all in the range of 0.69–0.96; correlations with Procrustes residuals in Supporting Information Dataset S5). Reflecting this high level of covariation, we found that different soil organism groups exhibited similar relationships with environmental variables, with MAT and soil pH consistently the most important drivers of community composition among different eukaryotic and prokaryotic groups (Supporting Information Figures S6–S8; Table S1).

Most organism groups exhibited low community turnover in space, with relatively high turnover apparent only among mycorrhizal fungal and animal communities. There was also weak differentiation of soil biota between biogeographical realms (for instance, samples from Afrotropic and Neotropic realms clustered together on ordination plots; Supporting Information Figure S6) and in relationship to historical biome continuity (whether or not the biome has existed undisturbed for >21 kyr; Supporting Information Table S1). The presence of potential ecosystem engineers, in the form of EcM fungi, also had weak to moderate effects on other soil organism groups (Supporting Information Table S1).

4 | DISCUSSION

Many issues related to large-scale pattern in biodiversity, ranging from environmental education to unravelling the consequences of global-scale ecological processes, are informed by biome descriptions, which provide a simplified representation of the structure of life on Earth. We used empirical soil biodiversity data and a statistical clustering approach to define soil microbiomes and show how these reflect the importance of edaphic variables, notably soil pH. The structure of soil microbiomes deviates importantly from existing terrestrial biome classifications, highlighting the drawbacks of using the latter as a surrogate for variation in soil microbial diversity, while also providing a practicable empirical alternative.

4.1 | Soil microbiomes in the context of existing terrestrial biomes

The soil microbiome types defined here show limited overlap with existing terrestrial biome classifications, which is consistent with emerging evidence of mismatches between the distribution patterns of above- and belowground organisms (Cameron et al., 2019). The proposed prokaryotic biome types are not paralleled by any existing biome (Mucina, 2019) or soil type classification (Hartemink et al., 2013), nor do they resemble the five bacterial ecological groups distinguished by Delgado-Baquerizo, Reich, Trivedi, et al. (2020: high pH, low pH, drylands, low plant productivity and dry-forest environments). The proposed microbial eukaryotic biome types also show little similarity to the biome system of Olson et al. (2001) (Figure 3; Supporting Information Figure S9). By contrast, similarity with the classifications of Whittaker (1975) and Walter (1976) is relatively high for biomes in warmer climates, but weaker for those in colder climates (Supporting Information Figure S9). Soil leaching and accompanying decreases in pH are more intense in warm and humid conditions (Slessarev et al., 2016), with the results particularly evident in older soils (Delgado-Baquerizo, Reich, Bardgett, et al., 2020). In cold environments, which include younger deglaciated soils, pH is expected to be more closely dependent on the mineralogy of the bedrock. Thus, precipitation, which partly defines Whittaker's and Walter's biomes, is likely to change in parallel with pH in warm environments (where higher precipitation is associated with lower pH) but less so in cold environments. In our dataset, the correlation between pH and precipitation was indeed stronger at warm (above median temperature; $r = -0.56$, $p < .001$) than cold (below median temperature; $r = -0.19$, $p = .01$) sites. Regional studies have shown a contrast between vegetation on acid and basic soils in subpolar and alpine (Lenoir et al., 2010; Virtanen et al., 2006), boreal (Chytrý et al., 2012; Noreika et al., 2019) and temperate (Axmanová et al., 2012; Chytrý et al., 2003) climatic zones, and Tedersoo, Anslan, et al. (2020) found analogous differences in fungal communities. The proposed microbial eukaryotic biome types reflect this; only in warmer areas do biomes separate along the precipitation gradient. Hence, using global terrestrial biomes as a framework for the study of prokaryotic biogeography or that of soil eukaryotes in higher latitudes appears to be of particularly low value.

The distribution of soil microbiomes reflects recorded relationships between soil communities and potential abiotic drivers, and earlier studies on this question. Soil pH and MAT were consistently among the strongest drivers of composition of different groups, corroborating earlier findings (Delgado-Baquerizo et al., 2018; Tedersoo et al., 2014; Větrovský et al., 2019), and the major groups of eukaryotes and prokaryotes in soil consistently

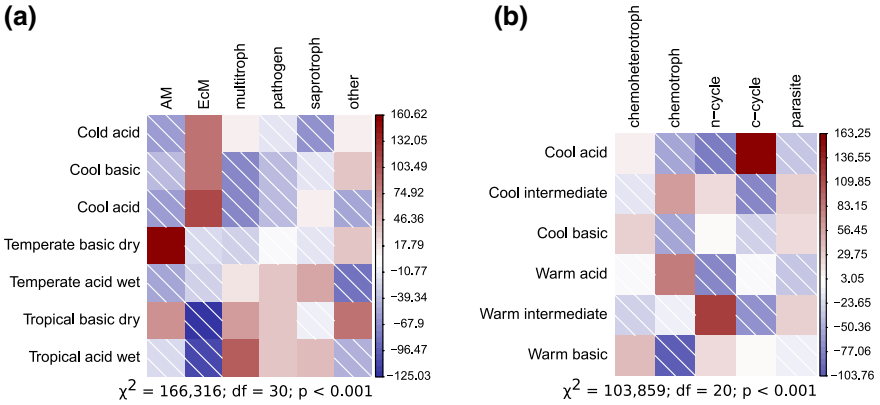

FIGURE 2 Relative contribution of different guilds and functional groups to clusters of microbial eukaryotic and prokaryotic diversity in soil. (a) Microbial eukaryote (ITS) clusters ($n = 7$). (b) Prokaryote (16S) clusters ($n = 6$). Each grid shows the relative magnitude of standardized Pearson residuals from χ^2 analysis (statistics shown below each panel) of a contingency table comprising read counts per guild in each cluster. Red cells indicate a positive association between guilds and clusters; blue hashed cells indicate negative associations. Colour intensity indicates the relative size of the residuals. The "other" category in panel (a) reflects non-fungal microbial eukaryotes

exhibited parallel variation at the global scale (see also local case studies of plant and soil microbial communities: Neuenkamp et al., 2018; Nottingham et al., 2018). Spatial constraints on distribution were most apparent among mycorrhizal fungi and soil animals, perhaps reflecting the necessity to associate with an appropriate host (mycorrhizal fungi; Tedersoo et al., 2020) or weak dispersal ability (soil animals; Wu et al., 2011). Meanwhile, the importance of unique regional history and the presence of ecosystem engineers, such as EcM trees (Tedersoo & Bahram, 2019), for soil biota appears relatively weak in comparison to other drivers. The present study focused on relatively natural ecosystems, hence the biotic communities emerging from pronounced interaction with anthropogenic activities (i.e., anthropomes; Ellis & Ramankutty, 2008) were not considered explicitly. It is likely that the relationships and emergent global patterns described here exhibit significant local variation in relationship to human activites, such as urbanization, agriculture and forestry (e.g., Carvalho et al., 2016; Moora et al., 2015; Schmidt et al., 2017).

Terrestrial biomes characterized by macro-organisms are clearly not predictive of patterns in soil microbial diversity and *vice versa*. Besides describing the ecology of the different organism groups involved, this might reflect the distinction between structure and diversity, which is mediated by organism functional and dispersal traits (Kivlin et al., 2014; Locey, 2010; Martiny et al., 2011). It is also important to note that, conceptually, biomes should not be considered single-level, simple community structures; rather, they are constructs of multiple levels of complexity, reflecting the hierarchy of evolutionary and ecological drivers. Mucina et al. (2021) have proposed the currently lowest-tier terrestrial biome classification, regional biomes (*sensu* Mucina et al., 2021), in order to represent precisely the joint effects of climatic and soil characteristics. We envisage that such a regional-scale comparison could yield a better fit between terrestrial biomes and soil microbiomes. Moreover, the disparity between the terrestrial biome structures derived on the basis of vegetation formations and the soil microbiomes does not preclude their parallel use. Each of these schemes, when applied appropriately, brings different insights into the patterning and dynamics of global biodiversity.

4.2 | Application of soil microbiomes

Incorporating a microbial component into biogeographical and macroecological frameworks that rely on defining large-scale biotic community structure is important for attempts to understand, predict and mitigate global change (Boit et al., 2016) and, in particular, those addressing changes in soil microbial diversity and functionality (Xu et al., 2020). For instance, soil microbiomes can inform sampling designs and provide explanatory variables for large-scale microbial ecology studies. Moreover, they would allow a microbial component to be added to global ecosystem models that are usually dominated by parameters pertinent to macrobiotic elements (e.g., plant and plant guilds, vegetation types). Furthermore, and perhaps surprisingly, the distinction between grassland and forest ecosystems (open vs. closed; one of the core tenets of any terrestrial biome scheme) does not emerge in the soil microbiome classification (Figure 3; Supporting Information Figures S4 and S9). As shown for R. H. Whittaker's biome classification, a large fraction of the temperature × precipitation factor space provides habitat suitable for both forest and grassland (Bond, 2005). Yet, routinely, forest and grassland biomes are considered separately; for example, by restricting sampling to one or the other, or routinely focusing on different microbial functions in each (Garland et al., 2021), based on the assumption that they represent coherent units with different features and relevance. It might well be that the characteristics of the microbial communities targeted by such studies in fact transcend the traditional forest and grassland biome dichotomy.

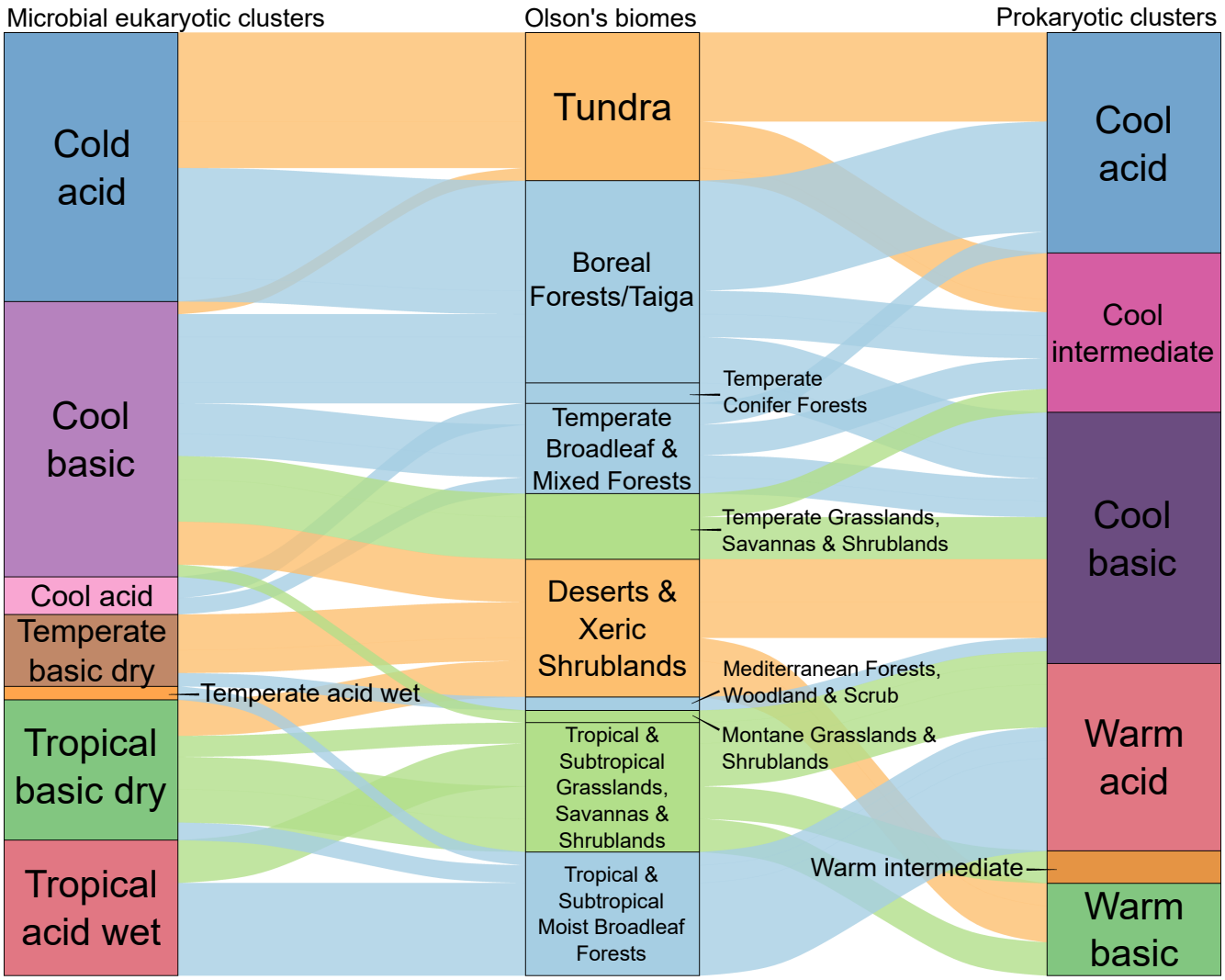

FIGURE 3 Comparison of proposed soil biomes with the biome classification of Olson et al. (2001). Flows represent the overlap among 10 Olson's biomes and the seven clusters of microbial eukaryotic diversity (left-hand classification) and six clusters of prokaryotic diversity (right-hand classification), on the basis of geographical coincidence in 0.5° × 0.5° interpolated map grid cells. The sizes of the clusters and flows are proportional to their abundance globally. Only flows representing coincidence in >500 grid cells are shown, thus also omitting four Olson's biomes. Individual Olson's biomes and flows from them to microbial eukaryotic and prokaryotic clusters are coloured according to the vegetation structure of the Olson's biome (green = grassland; blue = forest; orange = tundra/desert). Individual clusters in the eukaryotic and prokaryotic classifications are coloured according to the colour scheme used for clusters in Figure 1. Note that plants and soil animals (18S) are not included in the soil microbial eukaryotic clusters

4.3 | Conclusions

Focusing on the large-scale distribution of soil microbial communities can lead to important changes in our views about the global distribution of biological diversity. Correcting these views will help us to understand the natural world and to arrive at optimal decisions in response to the threats posed by global change. We advocate the use of empirical data to define soil microbiomes. However, we are aware that our study is only the beginning of the journey; a journey that should address the effects of spatial, temporal and taxonomic scales of sampling, identify the most informative environmental parameters and pattern-inference tools, and build large data archives.

ACKNOWLEDGMENTS

This study was supported by the European Regional Development Fund Centre of Excellence EcolChange (L.T., M.Z., all University of Tartu members), the University of Tartu (PLTOM20903 to M.Z.), the Estonian Research Council (MOBTP 105 to K.K.; PRG 609 to M.P.; PRG 1065 to M.M., C.G.B., J.D., J.O., Y.M. and T.V.; PRG 1170 to I.H.), the Russian Ministry of Science and Higher Education (# 075-15-2021-1396 to V.O.) and the Swedish Research Council Vetenskapsrådet (2017-05019 to M.B.). We thank the following for assistance with fieldwork: Sakeenah Binte Adenan, Talaat Ahmed, Asem A. Akhmetzhanova, Juha M. Alatalo, Elena Andriyanova, Amgaa Batbaatar, Charlotte Brown, James Cahill, Brenda B. Casper, Mikhail Cherosov, Saida Chideh, Ana P. Coelho, Matthew Coghill,

Sergey Dudov, Ezequiel Chimbioputo Fabiano, Nina Farish, Vladimir E. Fedosov, Sydney I. Glassman, Ruben Heleno, Aveliina Helm, Hugh A. L. Henry, Bruno Hérault, Indrek Hiiesalu, Teele Jairus, Rein Kalamees, Petr Kohout, Urmas Kõljalg, Darja Koltysheva, Lauri Laanisto, Ülo Mander, Mart Meriste, Dagmar Mucina, Jean-Pierre Munyampundu, Tanaka Muradzikwa, François-Xavier Naramabuye, Lena Neuenkamp, Ülo Niinemets, Casper Nyamukondiwa, Cherdchai Phosri, Sergei Põlme, Kersti Püssa, Ülle Reier, Argo Ronk, Alessandro Saitta, Olivia Semboli, Marina Semchenko, Alexey Seregin, Surya Sudheer, Clara P. Peña-Venegas, Claudia Paz, Yves Uwiragiye, Annelies J. Veraart and Anna Zobel.

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

Conceptualization: J.D., M.M., M.Ö., L.T., M.V. and M.Z.; Methodology: J.D., J.O., M.M., M.Ö., S.-K.S., L.T., M.V. and M.Z.; Investigation: all authors; Funding acquisition: L.T. and M.Z.; Project administration: L.T. and M.Z.; Writing – original draft: J.D. and M.Z.; Writing – review & editing: all authors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the Supporting Information or from the NCBI Sequence Read Archive (BioProject PRJNA659159).

ORCID

John Davison https://orcid.org/0000-0002-0161-6195
C. Guillermo Bueno https://orcid.org/0000-0002-7288-2271
Leho Tedersoo https://orcid.org/0000-0002-1635-1249

REFERENCES

Anslan, S., Bahram, M., Hiiesalu, I., & Tedersoo, L. (2017). PipeCraft: Flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. *Molecular Ecology Resources*, *17*, e234–e240. https://doi.org/10.1111/1755-0998.12692

Axmanová, I., Chytrý, M., Zelený, D., Li, C. F., Vymazalová, M., Danihelka, J., Horsák, M., Kočí, M., Kubešová, S., Lososová, Z., Otýpková, Z., Tichý, L., Martynenko, V. B., Baisheva, E. Z., Schuster, B., & Diekmann, M. (2012). The species richness–productivity relationship in the herb layer of European deciduous forests. *Global Ecology and Biogeography*, *21*, 657–667. https://doi.org/10.1111/j.1466-8238.2011.00707.x

Bastida, F., Eldridge, D. J., Abades, S., Alfaro, F. D., Gallardo, A., García-Velázquez, L., García, C., Hart, S. C., Pérez, C. A., Santos, F., Trivedi, P., Williams, M. A., & Delgado-Baquerizo, M. (2020). Climatic vulnerabilities and ecological preferences of soil invertebrates across biomes. *Molecular Ecology*, *29*, 752–761. https://doi.org/10.1111/mec.15299

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A. S., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y. J. K., Sanli, K., Eriksson, K. M., Vik, U., … Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, *4*, 914–919. https://doi.org/10.1111/2041-210X.12073

Boit, A., Sakschewski, B., Boysen, L., Cano-Crespo, A., Clement, J., Garcia-Alaniz, N., Kok, K., Kolb, M., Langerwisch, F., Rammig, A., Sachse, R., van Eupen, M., von Bloh, W., Clara Zemp, D., & Thonicke, K. (2016). Large-scale impact of climate change vs. land-use change on future biome shifts in Latin America. *Global Change Biology*, *22*, 3689–3701. https://doi.org/10.1111/gcb.13355

Bond, W. J. (2005). Large parts of the world are brown or black: A different view on the "Green World" hypothesis. *Journal of Vegetation Science*, *16*, 261–266.

Brooks, T. M., Mittermeier, R. A., Da Fonseca, G. A., Gerlach, J., Hoffmann, M., Lamoreux, J. F., Mittermeier, C. G., Pilgrim, J. D., & Rodrigues, A. S. L. (2006). Global biodiversity conservation priorities. *Science*, *313*, 58–61. https://doi.org/10.1126/science.1127609

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 1–9. https://doi.org/10.1186/1471-2105-10-421

Cameron, E. K., Martins, I. S., Lavelle, P., Mathieu, J., Tedersoo, L., Bahram, M., Gottschall, F., Guerra, C. A., Hines, J., Patoine, G., Siebert, J., Winter, M., Cesarz, S., Ferlian, O., Kreft, H., Lovejoy, T. E., Montanarella, L., Orgiazzi, A., Pereira, H. M., … Eisenhauer, N. (2019). Global mismatches in aboveground and belowground biodiversity. *Conservation Biology*, *33*, 1187–1192. https://doi.org/10.1111/cobi.13311

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the USA*, *108*, 4516–4522.

Castaño, C., Berlin, A., Brandström Durling, M., Ihrmark, K., Lindahl, B. D., Stenlid, J., Clemmensen, K. E., & Olsen, A. (2020). Optimized metabarcoding with Pacific Biosciences enables semi-quantitative analysis of fungal communities. *New Phytologist*, *228*, 1149–1158. https://doi.org/10.1111/nph.16731

Chytrý, M., Ermakov, N., Danihelka, J., Hájek, M., Hájková, P., Horsák, M., Kočí, M., Kubešová, S., Lustyk, P., Otýpková, Z., Pelánková, B., Valachovič, M., & Zelený, D. (2012). High species richness in hemiboreal forests of the northern Russian Altai, southern Siberia. *Journal of Vegetation Science*, *23*, 605–616. https://doi.org/10.1111/j.1654-1103.2011.01383.x

Chytrý, M., Tichý, L., & Roleček, J. (2003). Local and regional patterns of species richness in Central European vegetation types along the pH/calcium gradient. *Folia Geobotanica*, *38*, 429–442. https://doi.org/10.1007/BF02803250

Coleman, A. W. (2009). Is there a molecular key to the level of "biological species" in eukaryotes? A DNA guide. *Molecular Phylogenetics and Evolution*, *50*, 197–203. https://doi.org/10.1016/j.ympev.2008.10.008

Davison, J., Moora, M., Semchenko, M., Adenan, S. B., Ahmed, T., Akhmetzhanova, A. A., Alatalo, J. M., Al-Quraishy, S., Andriyanova, E., Anslan, S., Bahram, M., Batbaatar, A., Brown, C., Bueno, C. G., Cahill, J., Cantero, J. J., Casper, B. B., Cherosov, M., Chideh, S., … Öpik, M. (2021). Temperature and pH define the realised niche space of arbuscular mycorrhizal fungi. *New Phytologist*, *231*, 763–776. https://doi.org/10.1111/nph.17240

de Carvalho, T. S., Jesus, E. D., Barlow, J., Gardner, T. A., Soares, I. C., Tiedje, J. M., & Moreira, F. M. (2016). Land use intensification in the humid tropics increased both alpha and beta diversity of soil bacteria. *Ecology*, *97*, 2760–2771. https://doi.org/10.1002/ecy.1513

Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K., & Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science*, *359*, 320–325. https://doi.org/10.1126/science.aap9516

Delgado-Baquerizo, M., Reich, P. B., Bardgett, R. D., Eldridge, D. J., Lambers, H., Wardle, D. A., Reed, S. C., Plaza, C., Png, G. K., Neuhauser, S., Berhe, A. A., Hart, S. C., Hu, H. W., He, J. Z., Bastida, F., Abades, S., Alfaro, F. D., Cutler, N. A., Gallardo, A., … Fierer, N.

(2020). The influence of soil age on ecosystem structure and function across biomes. *Nature Communications*, *11*, 4721. https://doi.org/10.1038/s41467-020-18451-3

Delgado-Baquerizo, M., Reich, P. B., Trivedi, C., Eldridge, D. J., Abades, S., Alfaro, F. D., Bastida, F., Berhe, A. A., Cutler, N. A., Gallardo, A., García-Velázquez, L., Hart, S. C., Hayes, P. E., He, J. Z., Hseu, Z. Y., Hu, H. W., Kirchmair, M., Neuhauser, S., Pérez, C. A., … Singh, B. K. (2020). Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nature Ecology and Evolution*, *4*, 210–220. https://doi.org/10.1038/s41559-019-1084-y

Duda, P., Hayashi, Y., & Jaworski, M. (2012). On the strong convergence of the orthogonal series-type kernel regression neural networks in a non-stationary environment. In: L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh & J. M. Zurada (Eds), *International conference on artificial intelligence and soft computing* (pp. 47–54). Springer.

Edgar, R. C. (2018). Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics*, *34*, 2371–2375. https://doi.org/10.1093/bioinformatics/bty113

Ellis, E. C., & Ramankutty, N. (2008). Putting people in the map: Anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, *6*, 439–447. https://doi.org/10.1890/070062

Garland, G., Banerjee, S., Edlinger, A., Miranda Oliveira, E., Herzog, C., Wittwer, R., Philippot, L., Maestre, F. T., & van der Heijden, M. G. A. (2021). A closer look at the functions behind ecosystem multifunctionality: A review. *Journal of Ecology*, *109*, 600–613. https://doi.org/10.1111/1365-2745.13511

Guerin, G. R., Williams, K. J., Leitch, E., Lowe, A. J., & Sparrow, B. (2021). Using generalised dissimilarity modelling and targeted field surveys to gap-fill an ecosystem surveillance network. *Journal of Applied Ecology*, *58*, 766–776. https://doi.org/10.1111/1365-2664.13814

Guerra, C. A., Heintz-Buschart, A., Sikorski, J., Chatzinotas, A., Guerrero-Ramírez, N., Cesarz, S., Beaumelle, L., Rillig, M. C., Maestre, F. T., Delgado-Baquerizo, M., Buscot, F., Overmann, J., Patoine, G., Phillips, H. R. P., Winter, M., Wubet, T., Küsel, K., Bardgett, R. D., Cameron, E. K., … Eisenhauer, N. (2020). Blind spots in global soil biodiversity and ecosystem function research. *Nature Communications*, *11*, 3870. https://doi.org/10.1038/s41467-020-17688-2

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., … Christen, R. (2012). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*, D597–D604. https://doi.org/10.1093/nar/gks1160

Hartemink, A. E., Krasilnikov, P., & Bockheim, J. G. (2013). Soil maps of the world. *Geoderma*, *207*, 256–267. https://doi.org/10.1016/j.geoderma.2013.05.003

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics and Data Analysis*, *52*, 258–271. https://doi.org/10.1016/j.csda.2006.11.025

Jacobson, A. P., Riggio, J., Tait, A. M., & Baillie, J. E. (2019). Global areas of low human impact ('Low Impact Areas') and fragmentation of the natural world. *Scientific Reports*, *9*, 1–3. https://doi.org/10.1038/s41598-019-50558-6

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 1–20. https://doi.org/10.1038/sdata.2017.122

Kivlin, S. N., Winston, G. C., Goulden, M. L., & Treseder, K. K. (2014). Environmental filtering affects soil fungal community composition more than dispersal limitation at regional scales. *Fungal Ecology*, *12*, 14–25. https://doi.org/10.1016/j.funeco.2014.04.004

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G. W., Hartmann, M., Kirk, P. M., Kohout, P., … Larsson, K.-H. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, *22*, 5271–5277. https://doi.org/10.1111/mec.12481

Lenoir, J., Gégout, J. C., Guisan, A., Vittoz, P., Wohlgemuth, T., Zimmermann, N. E., Dullinger, S., Pauli, H., Willner, W., Grytnes, J. A., Virtanen, R., & Svenning, J.-C. (2010). Cross-scale analysis of the region effect on vascular plant species diversity in southern and northern European mountain ranges. *PLoS One*, *5*, e15734. https://doi.org/10.1371/journal.pone.0015734

Locey, K. J. (2010). Synthesizing traditional biogeography with microbial ecology: The importance of dormancy. *Journal of Biogeography*, *37*, 1835–1841. https://doi.org/10.1111/j.1365-2699.2010.02357.x

Louca, S., Parfrey, L. W., & Doebeli, M. (2016). Decoupling function and taxonomy in the global ocean microbiome. *Science*, *353*, 1272–1277. https://doi.org/10.1126/science.aaf4507

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Martiny, J. B., Eisen, J. A., Penn, K., Allison, S. D., & Horner-Devine, M. C. (2011). Drivers of bacterial β-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences of the USA*, *108*, 7850–7854.

McMurdie, P. J., & Holmes, S. (2014). Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, *10*, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Moncrieff, G. R., Bond, W. J., & Higgins, S. I. (2016). Revising the biome concept for understanding and predicting global change impacts. *Journal of Biogeography*, *43*, 863–873. https://doi.org/10.1111/jbi.12701

Moora, M., Davison, J., Öpik, M., Metsis, M., Saks, Ü., Jairus, T., Vasar, M., & Zobel, M. (2015). Anthropogenic land use shapes the composition and phylogenetic structure of soil arbuscular mycorrhizal fungal communities. *FEMS Microbiology Ecology*, *90*, 609–621. https://doi.org/10.1111/1574-6941.12420

Mucina, L. (2019). Biome: Evolution of a crucial ecological and biogeographical concept. *New Phytologist*, *222*, 97–114. https://doi.org/10.1111/nph.15609

Mucina, L., Lötter, M. C., Rutherford, M. C., van Niekerk, A., Macintyre, P. D., Tsakalos, J. L., Timberlake, J., Adams, J. B., Riddin, T., & McCarthy, L. K. (2021). Forest biomes of Southern Africa. *New Zealand Journal of Botany*. https://doi.org/10.1080/0028825X.2021.1960383

Neuenkamp, L., Moora, M., Öpik, M., Davison, J., Gerz, M., Männistö, M., Jairus, T., Vasar, M., & Zobel, M. (2018). The role of plant mycorrhizal type and status in modulating the relationship between plant and arbuscular mycorrhizal fungal communities. *New Phytologist*, *220*, 1236–1247. https://doi.org/10.1111/nph.14995

Nilsson, R. H., Taylor, A. F., Adams, R. I., Baschien, C., Bengtsson-Palme, J., Cangren, P., Coleine, C., Daniel, H. M., Glassman, S. I., Hirooka, Y., Laszlo, I., Iršėnaitė, R., Martin-Sanchez, P. M., Meyer, W., Oh, S. Y., Sampaio, J. P., Seifert, K., Sklenář, F., Dirk Stubbe, D. S., … Abarenkov, K. (2018). Taxonomic annotation of public fungal ITS sequences from the built environment – a report from an April 10–11, 2017 workshop (Aberdeen, UK). *MycoKeys*, *28*, 65–82.

Noreika, N., Helm, A., Öpik, M., Jairus, T., Vasar, M., Reier, Ü., Kook, E., Riibak, K., Kasari, L., Tullus, H., Tullus, T., Lutter, R., Oja, E., Saag, A., Randlane, T., & Pärtel, M. (2019). Forest biomass, soil and biodiversity relationships originate from biogeographic affinity and direct ecological effects. *Oikos*, *128*, 1653–1665. https://doi.org/10.1111/oik.06693

Nottingham, A. T., Fierer, N., Turner, B. L., Whitaker, J., Ostle, N. J., McNamara, N. P., Bardgett, R. D., Leff, J. W., Salinas, N., Silman, M. R., Kruuk, L. E. B., & Meir, P. (2018). Microbes follow Humboldt: Temperature drives plant and soil microbial diversity patterns from

the Amazon to the Andes. *Ecology*, *99*, 2455–2466. https://doi.org/10.1002/ecy.2482

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., & Wagner, H. (2020). *vegan: Community ecology package* (R package version 2.5-7). https://CRAN.R-project.org/package=vegan

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, *51*, 933–938.

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Pawlowski, J., Audic, S., Adl, S., Bass, D., Belbahri, L., Berney, C., Bowser, S. S., Cepicka, I., Decelle, J., Dunthorn, M., Fiore-Donno, A. M., Gile, G. H., Holzmann, M., Jahn, R., Jirků, M., Keeling, P. J., Kostka, M., Kudryavtsev, A., Lara, E., … de Vargas, C. (2012). CBOL protist working group: Barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biology*, *10*, e1001419. https://doi.org/10.1371/journal.pbio.1001419

Põlme, S., Abarenkov, K., Nilsson, R. H., Lindahl, B. D., Clemmensen, K. E., Kauserud, H., Nguyen, N., Kjøller, R., Bates, S. T., Baldrian, P., Frøslev, T. G., Adojaan, K., Vizzini, A., Suija, A., Pfister, D., Baral, H. O., Järv, H., Madrid, H., Nordén, J., … Tedersoo, L. (2020). FungalTraits: A user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Diversity*, *105*, 1–16. https://doi.org/10.1007/s13225-020-00466-2

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, *35*, 7188–7196. https://doi.org/10.1093/nar/gkm864

Ray, N., & Adams, J. (2001). A GIS-based vegetation map of the world at the last glacial maximum (25,000–15,000 BP). *Internet Archaeology*, *11*, 1–44. https://doi.org/10.11141/ia.11.2

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

Schliep, K., & Hechenbichler, K. (2016). *kknn: weighted k-nearest neighbors* (R package version 1.3.1). https://CRAN.R-project.org/package=kknn

Schmidt, D. J., Pouyat, R., Szlavecz, K., Setälä, H., Kotze, D. J., Yesilonis, I., Cilliers, S., Hornung, E., Dombos, M., & Yarwood, S. A. (2017). Urbanization erodes ectomycorrhizal fungal diversity and may cause microbial communities to converge. *Nature Ecology and Evolution*, *1*, 1–9. https://doi.org/10.1038/s41559-017-0123

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., & Chen, W. & Fungal Barcoding Consortium. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the USA*, *109*, 6241–6246.

Shen, C., Liang, W., Shi, Y., Lin, X., Zhang, H., Wu, X., Xie, G., Chain, P., Grogan, P., & Chu, H. (2014). Contrasting elevational diversity patterns between eukaryotic soil microbes and plants. *Ecology*, *95*, 3190–3202. https://doi.org/10.1890/14-0310.1

Slessarev, E. W., Lin, Y., Bingham, N. L., Johnson, J. E., Dai, Y., Schimel, J. P., & Chadwick, O. A. (2016). Water balance creates a threshold in soil pH at the global scale. *Nature*, *540*, 567–569. https://doi.org/10.1038/nature20139

Smith, J. R., Letten, A. D., Ke, P. J., Anderson, C. B., Hendershot, J. N., Dhami, M. K., Dlott, G. A., Grainger, T. N., Howard, M. E., Morrison, B. M. L., Routh, D., San Juan, P. A., Mooney, H. A., Mordecai, E. A., Crowther, T. W., & Daily, G. C. (2018). A global test of ecoregions. *Nature Ecology and Evolution*, *2*, 1889–1896. https://doi.org/10.1038/s41559-018-0709-x

Tedersoo, L., & Anslan, S. (2019). Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences. *Environmental Microbiology Reports*, *11*, 659–668. https://doi.org/10.1111/1758-2229.12776

Tedersoo, L., Anslan, S., Bahram, M., Drenkhan, R., Pritsch, K., Buegger, F., Padari, A., Hagh-Doust, N., Mikryukov, V., Gohar, D., Amiri, R., Hiiesalu, I., Lutter, R., Rosenvald, R., Rähn, E., Adamson, K., Drenkhan, T., Tullus, H., Jürimaa, K., … Abarenkov, K. (2020). Regional-scale in-depth analysis of soil fungal diversity reveals strong pH and plant species effects in Northern Europe. *Frontiers in Microbiology*, *11*, 1953. https://doi.org/10.3389/fmicb.2020.01953

Tedersoo, L., & Bahram, M. (2019). Mycorrhizal types differ in ecophysiology and alter plant nutrition and soil processes. *Biological Reviews*, *94*, 1857–1880. https://doi.org/10.1111/brv.12538

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios, A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., … Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science*, *346*, 1078–1087. https://doi.org/10.1126/science.1256688

Tedersoo, L., Bahram, M., & Zobel, M. (2020). How mycorrhizal associations drive plant population and community biology. *Science*, *367*, eaba1223. https://doi.org/10.1126/science.aba1223

Tedersoo, L., & Lindahl, B. (2016). Fungal identification biases in microbiome projects. *Environmental Microbiology Reports*, *8*, 774–779. https://doi.org/10.1111/1758-2229.12438

Tedersoo, L., Tooming-Klunderud, A., & Anslan, S. (2018). PacBio metabarcoding of Fungi and other eukaryotes: Errors, biases and perspectives. *New Phytologist*, *217*, 1370–1385. https://doi.org/10.1111/nph.14776

Tipton, L., Zahn, G., Datlof, E., Kivlin, S. N., Sheridan, P., Amend, A. S., & Hynson, N. A. (2019). Fungal aerobiota are not affected by time nor environment over a 13-y time series at the Mauna Loa Observatory. *Proceedings of the National Academy of Sciences of the USA*, *116*, 25728–25733.

Vasar, M., Davison, J., Neuenkamp, L., Sepp, S.-K., Young, J. P., Moora, M., & Öpik, M. (2021). User-friendly bioinformatics pipeline gDAT (graphical downstream analysis tool) for analysing rDNA sequences. *Molecular Ecology Resources*, *21*, 1380–1392. https://doi.org/10.1111/1755-0998.13340

Větrovský, T., Kohout, P., Kopecký, M., Machac, A., Man, M., Bahnmann, B. D., Brabcová, V., Choi, J., Meszárošová, L., Human, Z. R., Lepinay, C., Lladó, S., López-Mondéjar, R., Martinović, T., Mašínová, T., Morais, D., Navrátilová, D., Odriozola, I., Štursová, M., … Baldrian, P. (2019). A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nature Communications*, *10*, 1–9. https://doi.org/10.1038/s41467-019-13164-8

Virtanen, R., Oksanen, J., Oksanen, L., & Razzhivin, V. Y. (2006). Broad-scale vegetation-environment relationships in Eurasian high-latitude areas. *Journal of Vegetation Science*, *17*, 519–528. https://doi.org/10.1111/j.1654-1103.2006.tb02473.x

Walter, H. (1976). *Die Ökologischen Systeme der Kontinente (Biogeosphäre); Prinzipien ihrer Gliederung mit Beispielen*. Gerold.

Whittaker, R. H. (1975). *Communities and ecosystems*, 2nd ed. Macmillan.

Wijayawardene, N. N., Hyde, K. D., Al-Ani, L. K., Tedersoo, L., Haelewaters, D., Rajeshkumar, K. C., Zhao, R. L., Aptroot, A., Leontyev, D. V., Saxena, R. K., & Tokarev, Y. S. (2020). Outline of Fungi and fungus-like taxa. *Mycosphere*, *11*, 1060–1456. https://doi.org/10.5943/mycosphere/11/1/8

Wu, T., Ayres, E., Bardgett, R. D., Wall, D. H., & Garey, J. R. (2011). Molecular study of worldwide distribution and diversity of soil animals. *Proceedings of the National Academy of Sciences of the USA*, *108*, 17720–17725.

Xu, X., Wang, N., Lipson, D., Sinsabaugh, R., Schimel, J., He, L., Soudzilovskaia, N. A., & Tedersoo, L. (2020). Microbial macroecology: In search of mechanisms governing microbial biogeographic patterns. *Global Ecology and Biogeography*, *29*, 1870–1886. https://doi.org/10.1111/geb.13162

BIOSKETCH

Martti Vasar, **John Davison**, **Siim-Kaarel Sepp** and **Martin Zobel** are part of the Plant Ecology Lab at the University of Tartu. Their research tackles a range of questions relating to the ecology of plants and associated microbes. In recent years, they have conducted several large-scale studies to investigate the biogeography of arbuscular mycorrhizal fungi.

Ladislav Mucina is associated with the Harry Butler Institute of the Murdoch University in Perth, Western Australia. His research focuses on vegetation science, biodiversity science, restoration ecology and biogeography.

Leho Tedersoo is part of the Ecology of Biological Interactions group at the University of Tartu. He has led research into the global biogeography of fungi and other microbial groups. The research team includes a world-wide group of ecologists with a common interest in the large-scale diversity of soil micro-organisms.

SUPPORTING INFORMATION

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

How to cite this article: Vasar, M., Davison J., Sepp S.-K., Mucina L., Oja J., Al-Quraishy S., Anslan S., Bahram M., Bueno C. G., Cantero J. J., Decocq G., Fraser L., Hiiesalu I., Hozzein W. N., Koorem K., Meng Y., Moora M., Onipchenko V., Öpik M., … Zobel M. (2022). Global soil microbiomes: A new frontline of biome-ecology research. *Global Ecology and Biogeography*, *31*, 1120–1132. https://doi.org/10.1111/geb.13487