

# Metagenomic analysis of fungal assemblages at a regional scale in high-altitude temperate forest soils: alternative methods to determine diversity, composition and environmental drivers

Stephanie Hereira-Pacheco<sup>1,2</sup>, Itzel Arias-Del Razo<sup>2</sup>,  
Alejandra Miranda-Carrazco<sup>3,4</sup>, Luc Dendooven<sup>4</sup>,  
Arturo Estrada-Torres<sup>2</sup> and Yendi E. Navarro-Noya<sup>1</sup>

<sup>1</sup> Laboratorio de Interacciones Bióticas, Centro de Investigación en Ciencias Biológicas, Universidad Autónoma de Tlaxcala, San Felipe Ixtacuixtla, Tlaxcala, Mexico

<sup>2</sup> Estación Científica La Malinche, Centro Tlaxcala de Biología de la Conducta, Universidad Autónoma de Tlaxcala, Tlaxcala, Tlaxcala, Mexico

<sup>3</sup> Departamento de Ciencias Ambientales, Unidad Lerma, Universidad Autónoma Metropolitana, Lerma de Villada, Estado de México, Mexico

<sup>4</sup> Laboratorio de Ecología de Suelos, Departamento de Biotecnología y Bioingeniería, Cinvestav, Ciudad de México, Ciudad de México, Mexico

## ABSTRACT

**Background:** Understanding the diversity and distribution of fungal communities at a regional scale is important since fungi play a crucial role in ecosystem functioning. Our study used environmental metagenomics to determine fungal communities in mountainous forest soils in the central highlands of Mexico.

**Methods:** We used four different bioinformatic workflows to profile fungal assemblages, *i.e.*, Geneious+UNITE, single- and paired-end microbial community profiling (MiCoP), and Kraken2.

**Results:** The workflows yielded different results; one detected a higher abundance of ectomycorrhizal (EcM) and saprophytic fungi, while the other identified more saprophytic and pathogenic fungi. Environmental, vegetation, and geographical factors determined the spatial distribution of soil fungi at a regional scale. Potential hydrogen (pH), calcium (Ca), magnesium (Mg), and silt content were detected as common drivers of fungal communities across different datasets enriched towards a functional guild. Vegetation traits were found to be more influential in shaping symbiotic fungi composition than saprotrophic and pathogenic fungi. This highlights the importance of considering vegetation traits when studying fungal community diversity and distribution. Clustering patterns of sampling points near the volcanoes indicated shared environmental and vegetation characteristics. A weak but significant distance decay in taxonomic similarity revealed that dispersal limitation contributed to fungal community composition, although it was not the primary factor in this study. Overall, this study provides important insights into the challenges and opportunities of studying fungal communities at a regional scale using metagenomic data.

Submitted 30 May 2024  
Accepted 25 September 2024  
Published 2 January 2025

Corresponding author  
Yendi E. Navarro-Noya,  
yendiebenezer.navarro.n@uatx.mx

Academic editor  
Xin Zhang

Additional Information and  
Declarations can be found on  
page 22

DOI 10.7717/peerj.18323

© Copyright  
2025 Hereira-Pacheco et al.

Distributed under  
Creative Commons CC-BY 4.0

## OPEN ACCESS

**Subjects** Bioinformatics, Ecology, Ecosystem Science, Molecular Biology, Mycology  
**Keywords** Bioinformatic workflows, Ecological omics, Environmental DNA, Forest microbiome, Soil mycobiome

## INTRODUCTION

Soil fungi play a crucial role in providing various ecosystem services within forests. They contribute to nutrient recycling, enhance soil stability, facilitate soil carbon storage, improve soil water retention, and mediate biological interactions by controlling or producing plant and insect diseases. Additionally, fungi establish symbiotic associations with a wide range of organisms, including vascular plants, algae, cyanobacteria, and others. They also serve as habitats and food sources for numerous soil-dwelling organisms (Bennett, 1998; Bhantana *et al.*, 2021; Ward *et al.*, 2022).

Despite the importance of soil fungi, our understanding of their spatial patterns and the ecological factors shaping their communities remains context-specific, with limited generalizability across ecosystems and at larger scales. This requires further exploration, as a deeper understanding of these ecological factors could help preserve ecosystems in the face of environmental changes. Traditional methods, such as sampling fruiting bodies, have extensively studied diversity of fungal communities (Peay & Bruns, 2014). However, the use of environmental DNA (eDNA) analysis and high-throughput sequencing (HTS) in recent years has enabled the recovery of more comprehensive information about microbial diversity and composition (Tedesco *et al.*, 2022). This new approach shows great promise in enhancing our understanding of fungi in much greater detail.

Soil fungi exhibited different assembly patterns depending on the geographic scale. Dispersal limitation, land use, climate, seasonality, edaphic factors, and plant community were identified as the main drivers of the fungal distribution (Glassman, Wang & Bruns, 2017; Lan *et al.*, 2022; Li *et al.*, 2022; Rasmussen *et al.*, 2018; Tedesco *et al.*, 2020; Vasco-Palacios *et al.*, 2020; Xing *et al.*, 2021; Zhao *et al.*, 2019). Typically, climate, historical dispersal limitations, and the covariation of climate with vegetation and edaphic properties have a limited effect on the fungal distribution at a regional geographic scale, as these factors are intrinsic to larger geographic scales (Lomolino, Riddle & Brown, 2006; Tedesco *et al.*, 2020). However, different patterns might emerge in mountainous ecosystems, especially in highly biodiverse areas, such as the Trans-Mexican Volcanic Belt (TMVB) region. The TMVB is composed of the highest mountains (mostly volcanoes) in Mexico, and it is part of the Mexican Transition Zone. In this biogeographical zone, the neotropical and nearctic biotas overlap (Arriaga-Jiménez, Rös & Halffter, 2018). Additionally, it is a center of diversification for several species (Moctezuma, Halffter & Escobar, 2016; Ruiz-Sánchez & Specht, 2013). Specific environmental factors contribute to the high level of endemisms found in the TMVB, exceeding predictions based solely on general macroecological factors (Rahbek *et al.*, 2006). The small surface area of these high mountain regions and their geographic isolation results in small and fragmented patches of vegetation (Mastretta-Yanes *et al.*, 2018). Furthermore, the complex geological and volcanic history in the region made the edaphic characteristics very diverse (Rodríguez *et al.*, 2019).

Soil fungal species and functional guilds respond differently to variations in soil characteristics, nutrient availability, and vegetation (Peay, Baraloto & Fine, 2013; Bahram *et al.*, 2020; Sui *et al.*, 2022). It is known that arbuscular mycorrhizal fungal (AMF) community composition is affected by abiotic variables (potential hydrogen (pH), rainfall, soil type, and land use), but not by geographical distance (Hazard *et al.*, 2013; Xu *et al.*, 2017). Both ectomycorrhizal (EcM) and saprotroph fungi are shaped by vegetation and forest type due to the presence of hosts, differences in litter composition or quality, and soil water-holding capacity (WHC) (Peay, Baraloto & Fine, 2013). Across different regions, the proportion of EcM plants had a strong effect on symbiotic fungi diversity in forests (Hui *et al.*, 2017; Kernaghan *et al.*, 2003; Tedersoo *et al.*, 2020), and in some regions on saprotrophic fungi (Khalid *et al.*, 2024). Additionally, the dispersal capacity can be different within functional groups of fungi. For example, EcM fungi have a high dispersal capacity due to their ability to disperse through wind-borne spores and animal vectors. In contrast, most AMF species are hypogeous, and while approximately one-third of the described species have the potential for aerial dispersal, some have large single-cell spores that limit long-distance dispersion (Chaudhary *et al.*, 2020).

Although many studies focus on the patterns and drivers of the diversity of fungal groups, such as EcM (Tedersoo, May & Smith, 2010) and AMF (Bouffaud *et al.*, 2016), only a few studies have investigated all fungi, and none have used a metagenomic approach on a regional scale. Metagenomic approaches have the advantage of not including a PCR step, which can introduce bias toward certain taxa. This study used HTS from eDNA and four bioinformatics workflows to determine fungal communities. The selected workflows were based on (1) the use of fungal-genome reference databases to map the reads, or (2) the mapping of reads against a genetic marker (barcode). Additionally, we selected methods that provide frequency or relative abundance data necessary for diversity analysis. The workflows yielded different results; one detected a higher abundance of ectomycorrhizal (EcM) and saprophytic fungi, while the other identified more saprophytic and pathogenic fungi. We used the results from the four workflows and determined how geographic distance, edaphic factors and nutrient content (environmental distance), and tree traits (vegetation) drove soil fungal community composition and diversity using datasets enriched towards functional guild at a regional scale in forest fragments dominated by two types of trees (deciduous and coniferous) in the east part of the TMVB in Mexico. We hypothesized that: (1) the spatial composition of EcM fungi was mainly affected by vegetation type and its characteristics while that of saprophytic fungi by the edaphic variables; and (2) the taxonomic composition of the database enriched with symbiotrophs would be best explained by tree composition and their traits, while when using the database enriched with saprotrophs, the taxonomic composition would be best explained by the edaphic factors and nutrient content.

Mountain forests in the neotropics are highly valuable ecosystems for their biodiversity and endemisms (Salinas *et al.*, 2021). However, they are vulnerable to climate change, and many worldwide are under great pressure due to land use change, fragmentation, and disturbance of the natural vegetation (Newbold *et al.*, 2020). More knowledge of the diversity and distribution of the species in this region is urgently needed to identify areas of

species richness (hotspots) and to designate protected areas to conserve their diversity. Additionally, this study explored various bioinformatic approaches for analyzing fungal communities using metagenomic data, which could aid in making informed decisions about bioinformatic workflows based on our findings.

## MATERIALS AND METHODS

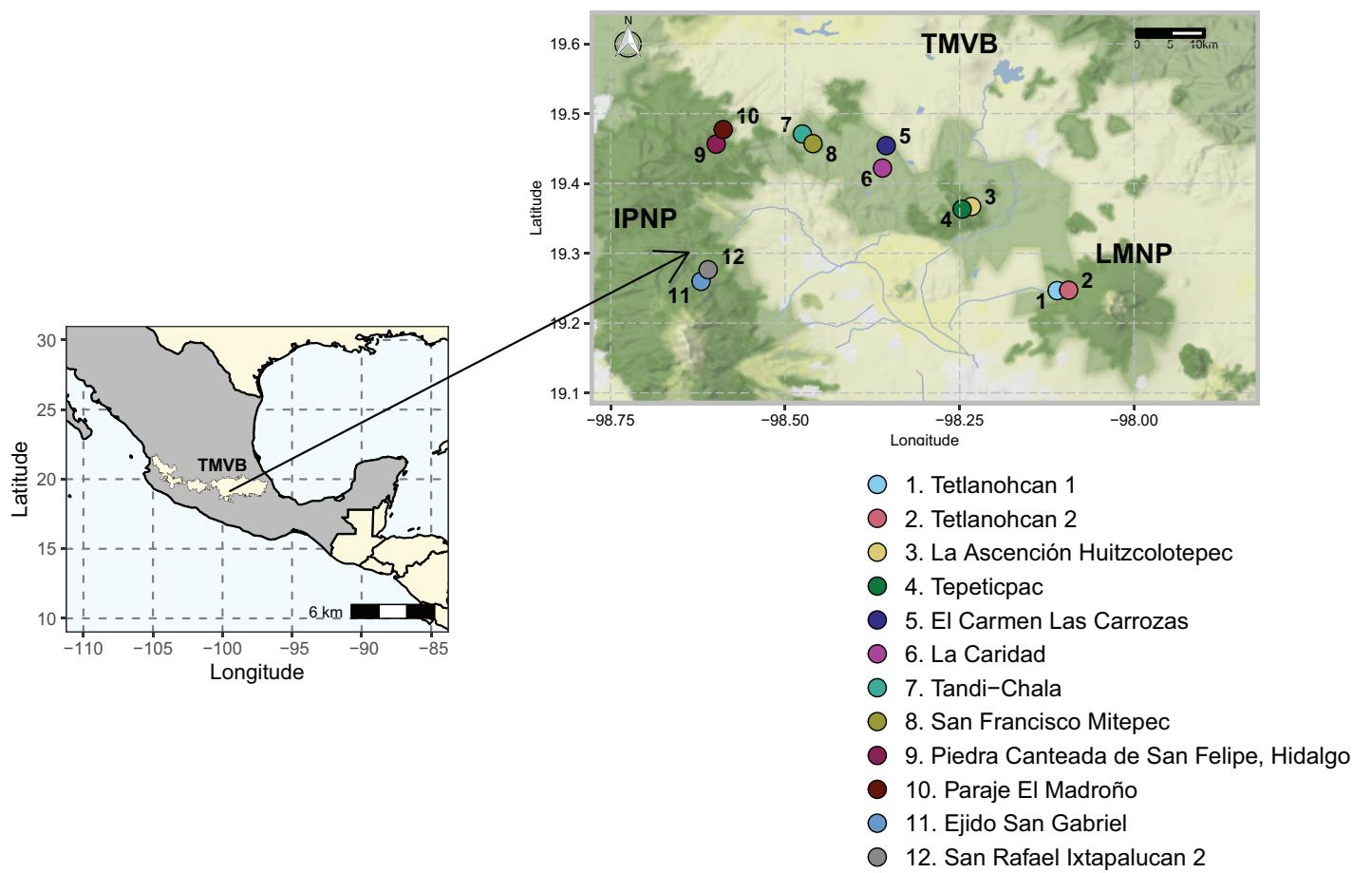
### Study site and soil sampling

The study area was located in the states of Tlaxcala and Puebla ( $N\ 19^{\circ}14'44.6821''\ W\ 98^{\circ}06'33.9509''-N\ 19^{\circ}16'36.7785''\ W\ 98^{\circ}36'36.0215''$ ). It included two National Parks, La Malinche (LMNP) and Iztaccíhuatl-Popocatépetl (IPNP), that are immersed in the most populated region of the continent (megalopolis), east of the TMVB. This region has a temperate subhumid climate. The average annual temperature is  $14\ ^{\circ}\text{C}$ , the average maximum temperature is around  $25\ ^{\circ}\text{C}$  and occurs in April and May, and the average minimum temperature is  $1.5\ ^{\circ}\text{C}$  in January. The average precipitation is 720 mm annually; with most rains in summer from June to September ([INEGI, 2024](#)). The elevation varies between 1,711 and 5,426 m above sea level (a.s.l.), and the original vegetation is temperate and subpolar coniferous forest, temperate and subpolar scrub, deciduous broadleaf forest, and temperate grasslands ([CONABIO, 2023](#)).

Sampling was done during the dry season of 2021. A total of 12 sampling sites distributed in six polygons ( $\approx 13\ \text{km}^2$ ), two sites per polygon were considered ([Fig. 1](#), [Table S1](#)). Sampling areas were located within forest fragments dominated by *Abies*, *Pinus*, and *Quercus* genera species. At each sampling site, three transects of 50 m were defined, and a composite sample of soil was collected from each transect by taking 400 g of soil every 10 m along the entire length. The O horizon was removed, and the underlying 0–10 cm topsoil was sampled with a small hand spade. All samples from each transect were pooled and 5 mm sieved to create a single composite sample. As such, 36 soil samples ( $N = 36$ ) were collected ( $6\ \text{polygons} \times 2\ \text{sites} \times 3\ \text{transects}$ ), and this field-based replication was maintained in the laboratory. Geographical coordinates were recorded at the beginning, middle, and endpoint of each transect. All soil samples were transported to the laboratory on ice and arrived within 60–90 min. Upon arrival, a sub-sample of 50 g was stored at  $-20\ ^{\circ}\text{C}$  until extracted for DNA as described below, while the rest was air-dried and characterized.

### Soil characterization

The WHC was measured by adding three 20-g soil sub-samples separately to a plastic funnel with Whatman No. 42 filter paper, applying 50 ml water, and letting them drain overnight ([Cassel & Nielsen, 1986](#)). At the same time, another sub-sample of 20 g was used to determine the moisture content. The soil was weighed, and the water retained, including previously contained water, was considered the WHC. Soil pH and electrolytic conductivity (EC) were measured in 1:2 soil-H<sub>2</sub>O suspension using a potentiometer/electrode (Hanna HI 2210; Hanna Instruments, Woonsocket, RI, USA) ([Thomas et al., 1996](#)). The nutrient concentrations were measured as follows: phosphorus (P) was measured in a photometer (Genesys<sup>TM</sup> 10S UV-VIS; Thermo Scientific, Waltham,



**Figure 1** Study area and 12 sampling sites between the National Parks *L. Malinche* (LMNP) and *Iztaccíhuatl-Popocatépetl* (IPNP) located on the east side of the Trans-Mexican Volcanic Belt (TMVB). Map from Stadia Maps ([stadiamaps.com](http://stadiamaps.com)).

[Full-size](#) DOI: [10.7717/peerj.18323/fig-1](https://doi.org/10.7717/peerj.18323/fig-1)

MA, USA), potassium (K) in a flame emission spectrophotometer (Cole-Parmer 2655-00 digital flame analyzer; Cole-Parmer Instrument Co., IL, USA), magnesium (Mg), calcium (Ca), iron (Fe), copper (Cu) and manganese (Mn) with an atomic absorption spectrometer (model SAVANTAA; GBC Scientific, Melbourne, Australia). Total nitrogen (N) was measured by the Kjeldahl method ([Bremner, 1996](#)), and the soil particle size distribution by the hydrometer method as described by [Gee & Bauder \(1986\)](#). Soil organic matter (SOM) was determined using the chromic acid titration method described by [Walkley & Black \(1934\)](#).

### Characterization of the vegetation

The vegetation was characterized at each sampling site. A 50 × 25 m quadrant was delineated within each transect, and individual trees were identified at the species level. Total height was measured using a forest clinometer. Coverage was determined using the crown diameter, and basal area was computed based on the diameter at breast height for each tree. Proportions were obtained by genus and type of tree (*i.e.*, conifer or

broadleaf). Means of measured characteristics were calculated by sampling sites, tree genus, and tree type.

### Soil DNA extraction and high-throughput sequencing

Metagenomic DNA was extracted using the DNeasy PowerSoil Pro Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Following the manufacturer's instructions, library construction was done with the Kapa Biosystems (Roche, CA, USA) kit. The size-selected libraries were amplified with three cycles of PCR and run on a Fragment Analyzer (AATTI, Ankeny, IA, USA) to confirm the absence of free primers and adaptor dimers, and the presence of DNA of the expected size range. Libraries were pooled in equimolar concentration and quantitated by qPCR on a Bio-Rad CFX Connect Real-Time System (Bio-Rad Laboratories Inc., Hercules, CA, USA). Sequencing was done in an Illumina Novaseq 6000 platform, using a 150 bp  $\times$  2 run, resulting in  $\sim$ 30 M reads per sample. Library preparation and sequencing were done by the "Centre d'expertise et de services Génome Québec" (Quebec, Canada).

### Sequence processing and taxonomic analysis

First, all reads were filtered to remove human sequences using BBMAP v.38.18 ([Bushnell, 2014](#)). Then, a quality score filter was applied using TRIMOMATIC v.0.39 to remove low-quality reads and/or those containing Illumina adaptors using the following parameters: ILLUMINACLIP:2:30:10; SLIDINGWINDOW:4:20; and MINLEN:30 ([Bolger, Lohse & Usadel, 2014](#)). The resulting high-quality sequences were analyzed using various bioinformatic workflows to determine fungal taxonomic composition. Three read-based pipelines were employed: Kraken2, MiCoP (Microbial Community Profiling), and Geneious+UNITE ([Fig. 2](#)). The Kraken2 workflow ([Wood, Lu & Langmead, 2019](#)) used a k-mer based approach and a machine learning classification algorithm against the PlusPFP index database (<https://benlangmead.github.io/aws-indexes/k2>) that contains fungi taxa constructed within the Kraken2 software. The abundance of taxonomic groups was determined with Bracken v.2.5 ([Lu et al., 2017](#)). Within the MiCoP workflow (repository cloned in November 2021; [LaPierre et al., 2019](#)) two options were used: single-end and paired-end. First, the reads whether single or paired, were mapped to a reference database of complete fungal genomes using the BWA-MEM v.0.6 software ([Li, 2013](#)). Subsequently, a two-stage process was used to classify the reads. Initially, all uniquely mapped reads were classified using MiCoP's built-in database and the abundance of each genome was determined in each sample based on these reads. Multi-mapped reads were assigned probabilistically to one of the genomes with probabilities proportional to the abundance obtained from uniquely-mapped reads of those genomes ([LaPierre et al., 2019](#)). The Geneious+UNITE workflow ([Bolyen et al., 2019](#)) was followed as suggested by [Smith et al. \(2020\)](#). First, reads were mapped against a modified UNITE (dynamic version 9, <http://unite.ut.ee/>) reference database ([Nilsson et al., 2019](#)) using the Geneious read mapper with 'Medium-Low Sensitivity/Fast sensitivity' settings and iterated twice ([Kearse et al., 2012](#)). The mapped reads were imported into the QIIME2 v.2022.2, dereplicated using q2-vsearch ([Rognes et al., 2016](#)), and filtered with 'qiime quality-control exclude-seqs' against the

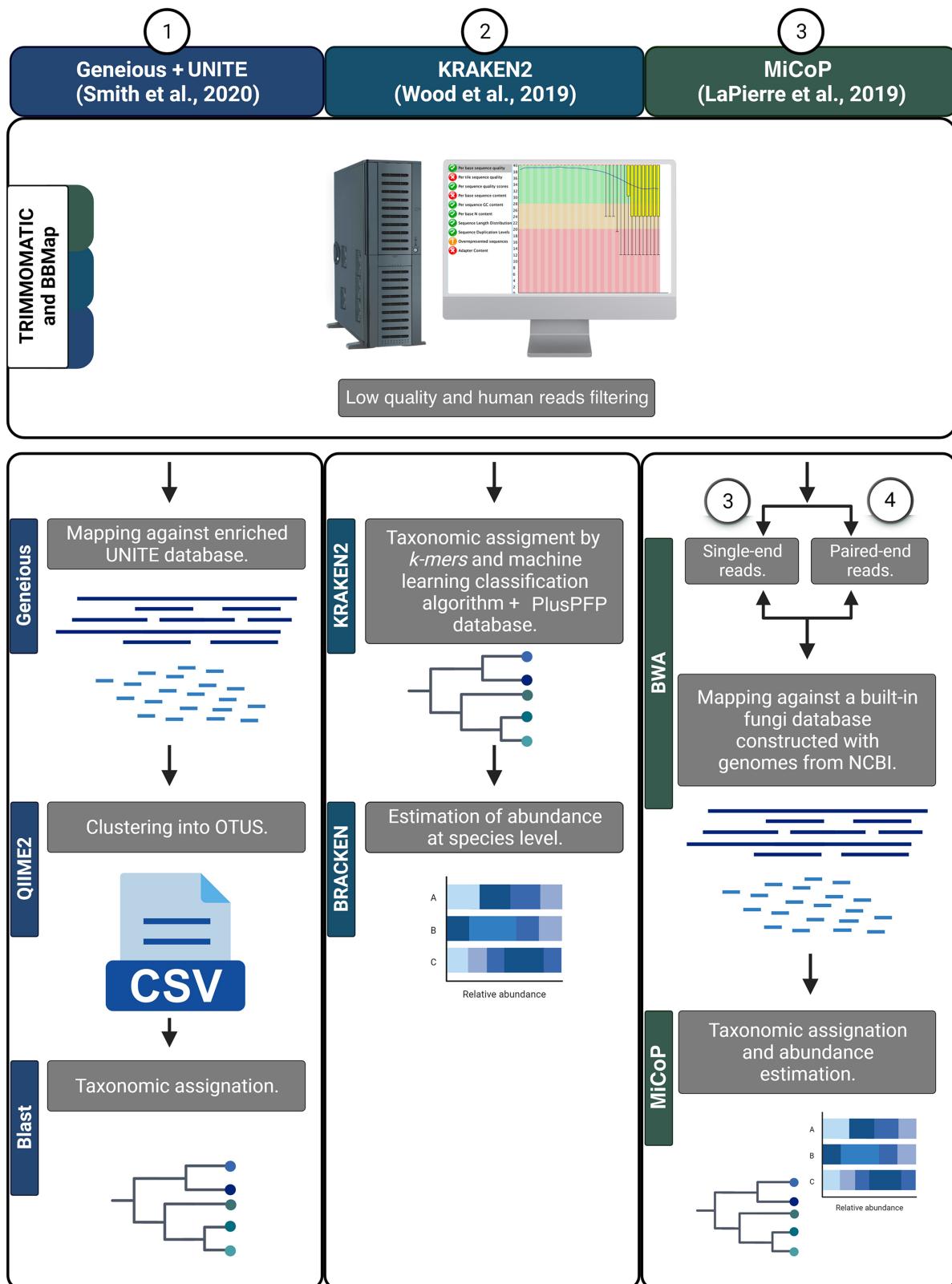


Figure 2 Bioinformatic workflows to determine fungal taxonomic profiles.

Full-size DOI: 10.7717/peerj.18323/fig-2

UNITE database. The remaining sequences were clustered *de novo* into OTUs at a 97% similarity using q2-vsearch ([Rognes et al., 2016](#); [Smith et al., 2020](#)). Taxonomic assignment was done with BLAST at 98% of identity, and unassigned reads were filtered out.

As the taxonomic unit obtained by the various pipelines was either taxonomic species or OTUs, we will refer to them as ‘features’. Tables displaying the frequency counts generated from each workflow will be referred to as ‘feature-tables’. To infer and annotate the ecological guilds of the feature-tables derived from the four workflows, we used the FUNGuild v.1.0 software and enriched the annotation manually ([Nguyen et al., 2016](#)).

## Statistical analysis

All statistical analyses were done using R v.4.3.1 ([R Core Team, 2023](#)), and figure generation with packages ggplot2 (v.3.4.4) and ggpubr (v.0.6.0) ([Wickham, 2011](#); [Kassambara, 2020](#)). To create maps, we used the ggmap package within the Stadia Maps using a personal API key ([Kahle & Wickham, 2013](#)). Heatmaps representing the relative abundance of taxonomic annotations were done with the ComplexHeatmap package v.2.14.0 ([Gu, Eils & Schlesner, 2016](#)). All figures were created using colorblind-friendly palettes from the R package rcartocolor v.2.1.1 ([Nowosad, 2023](#)).

Alpha diversity indexes were calculated using Hill numbers. This approach provides intuitive measures in the “effective number of features”, which facilitate interpretation and comparison. Hill numbers modulate the weight of species abundance with the parameter  $q$  (diversity order), *i.e.*,  $q = 0$  corresponds to the total number of features (richness),  $q = 1$  corresponds to frequent features (equivalent to Shannon exponential entropy), and  $q = 2$  to dominant features (equivalent to the reciprocal of the Simpson measure) ([Chao, Chiu & Jost, 2014](#)). The Hill numbers were calculated using the hillR package v.0.5.1 ([Li, 2018](#)). Kruskal-Wallis tests were performed to examine differences between sites with the ‘kruskal.test’ function within R (stats default package).

To compare taxonomic compositions within the Hill numbers framework, dissimilarity community matrices were computed from each workflow’s feature-tables at order  $q = 1$  (accounting for frequent features and equivalent to Classic Horn dissimilarity index; [Chao, Chiu & Jost, 2014](#)) using the function ‘hill\_taxa\_parti\_pairwise’ of the hillR package. Bray-Curtis dissimilarity matrices were calculated within phyloseq as this metric is widely used in similar ecological studies ([McMurdie & Holmes, 2013](#)). The similarity was calculated when necessary by subtracting the dissimilarity value from one (1-dissimilarity). A permutational multivariate analysis of variance (perMANOVA) was done with the adonis2 function ( $n = 999$ ) to test community differences between sites ( $n = 12$ ), and a permutational multivariate homogeneity of group dispersion analysis (PERMDISP) with the betadisper function. Both analyses were done with the vegan package v.2.5.3 ([Oksanen et al., 2017](#)). Additionally, a principal coordinate analysis (PCoA) was plotted to explore dissimilarity patterns of fungal communities obtained with each pipeline with the ‘pcoa’ function from ape v.5.5 ([Paradis, Claude & Strimmer, 2004](#)).

To determine the effect of the geographic distance on fungal communities, a geographical distance matrix was obtained using geographical coordinates with the function ‘distm’ from the geosphere package v.1.5-14 ([Hijmans, 2021](#)) (Table S1).

**Table 1** Summary of the frequencies obtained with different bioinformatic workflows.

Bioinformatic workflow	Total frequencies	Number of features <sup>a</sup>	Minimum frequency by sample	Maximum frequency by sample	Mean frequency by sample	Median frequency by sample	Mean Good's coverage (%)
Geneious+ UNITE	34,739	3,138	395	2,059	965	848	79.83
Single MiCoP	93,589	210	1,194	5,043	2,600	2,392	99.32
Paired MiCoP	1,281,436	234	21,654	52,713	35,595	35,008	99.94
Kraken2	1,021,469	87	17,680	42,316	28,374	27,185	100

Note:

<sup>a</sup> Taxonomic species (single and paired MiCoP, and Kraken2) or operational taxonomic unit at 97% of similarity (Geneious+UNITE).

Dispersal limitation was measured as an increasing dissimilarity in composition with geographic distance, *i.e.*, distance decay (Kivlin *et al.*, 2014). Mantel tests with the Spearman method were used to evaluate the association between community similarity and geographic distance using the 'cor.test' function and linear regressions with the 'lm' function (stats default R package). Partial Mantel tests were applied to assess correlations between environmental and vegetation matrices (Euclidean distances) and community matrices conditioned on spatial distances.

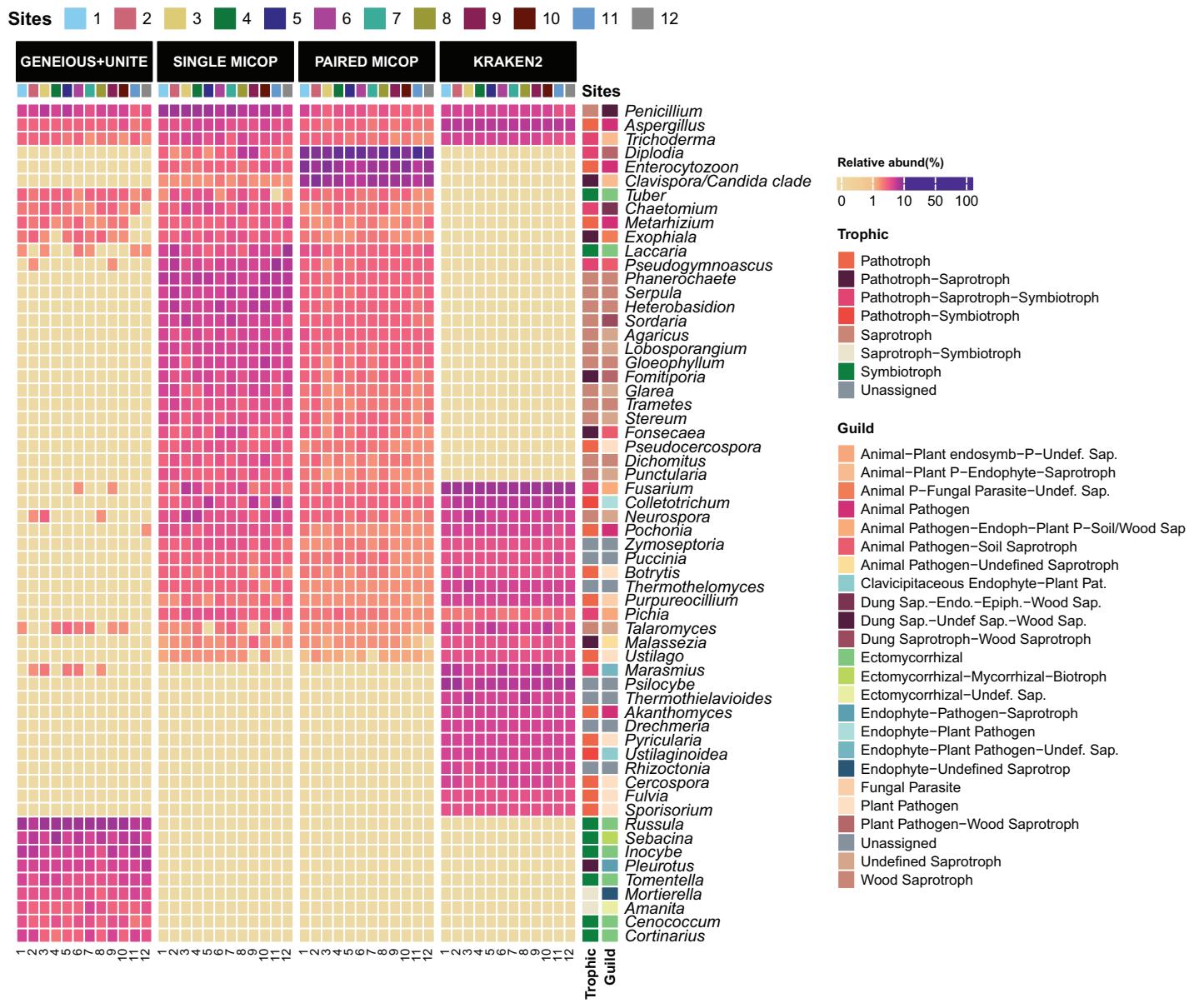
The following steps were taken to determine the effect of environmental factors (soil characteristics and nutrient content) and vegetation on fungal taxonomic compositions. First, variables were standardized (proportion data were log normalized, centered, and scaled). Next, collinearity among the variables was assessed using correlation plots generated with the corrplot package v.0.92. Variables exhibiting high correlation (Pearson correlation:  $|r| > 0.7$ ) were subsequently removed from the analysis (Gao *et al.*, 2020; Wei & Simko, 2021). To select the variables explaining most of the variance of fungal communities: (1) feature-tables were Hellinger-transformed, (2) a constrained correspondence analysis (CCA) was done with the transformed data followed by a fit of environmental factors onto an ordination with the function 'envfit', and (3) a redundancy analysis (RDA) was done with the transformed data followed by a forward selection using 'capscale' function with the Horn and Bray-Curtis distances that account for abundance. Both CCA and RDA were done with 999 permutations and variables retained at  $p < 0.05$ . All ecological analyses were done with the vegan package.

Raw sequences are available at the NCBI Sequence Read Archive under BioProject PRJNA875082. All scripts and metadata for statistical analyses and visualizations are available at <https://zenodo.org/doi/10.5281/zenodo.13629859>.

## RESULTS

### Alpha diversity patterns of the fungal communities obtained by different workflows in the soils across sites

After implementing the described workflows, the total number of reads varied markedly (Table 1). The strategy that recovered the most reads and/or total frequency was Paired MiCoP, with 1,281,436 reads, and Geneious+UNITE, the least, with 34,739 counts. The

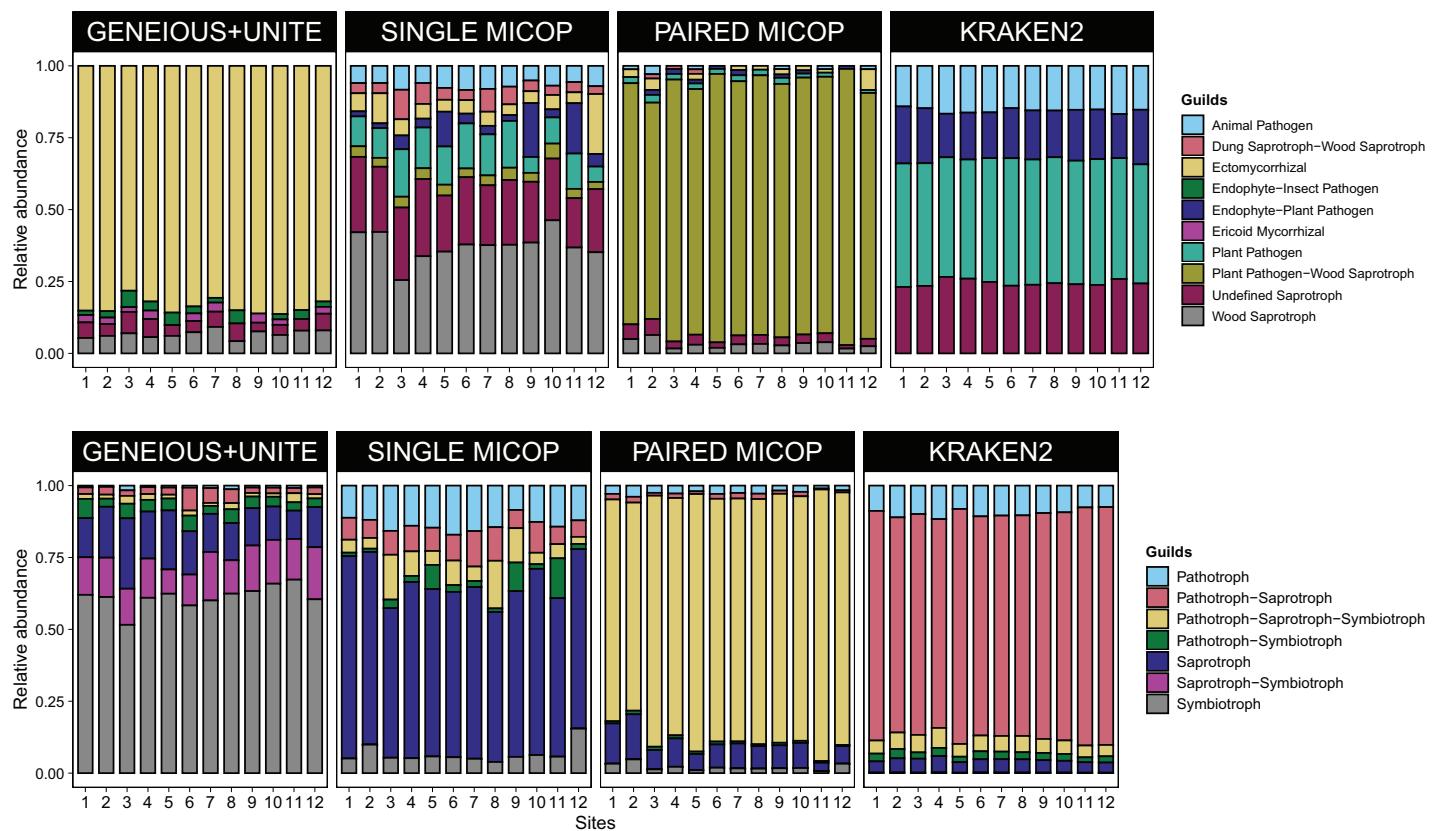


**Figure 3** Heatmap of the relative abundance of the 50 most abundant genera across the sampling sites (1–12) obtained by different bioinformatic workflows, i.e., GENEIOUS+UNITE, SINGLE and PAIRED MICOP and KRAKEN2 (see Fig. 2). Annotations on the right indicate the guild and trophic assignment with FUNGuild. Colored numbers refer to the different sampling sites and their location is given in Fig. 1.

Full-size DOI: 10.7717/peerj.18323/fig-3

number of features detected varied across the different workflows. Geneious+UNITE identified the highest number of features (3,138) and Kraken2 the least, with 87. The mean Good's coverage by workflow ranged from 79% to 100% of the total diversity.

The different workflows gave contrasting results that affected the calculation of the alpha diversity. The workflows, except Kraken2, which had a low variation, coincided in sites 11 and 12 (coniferous forest located in the IPNP), which had the lowest richness ( $q = 0$ ) (Fig. S1). Diversity values, Hill numbers at  $q = 1$  and 2 (frequent and dominant



**Figure 4** Bar plots of the relative abundance of the 20 most abundant in (A) guilds and (B) trophic groups across different sites (1–12) and workflows, i.e., GENEIOUS+UNITE, SINGLE and PAIRED MiCoP and KRAKEN2 (see Fig. 2). Numbers in the x-axis refer to the different sampling sites and their location is given in Fig. 1.

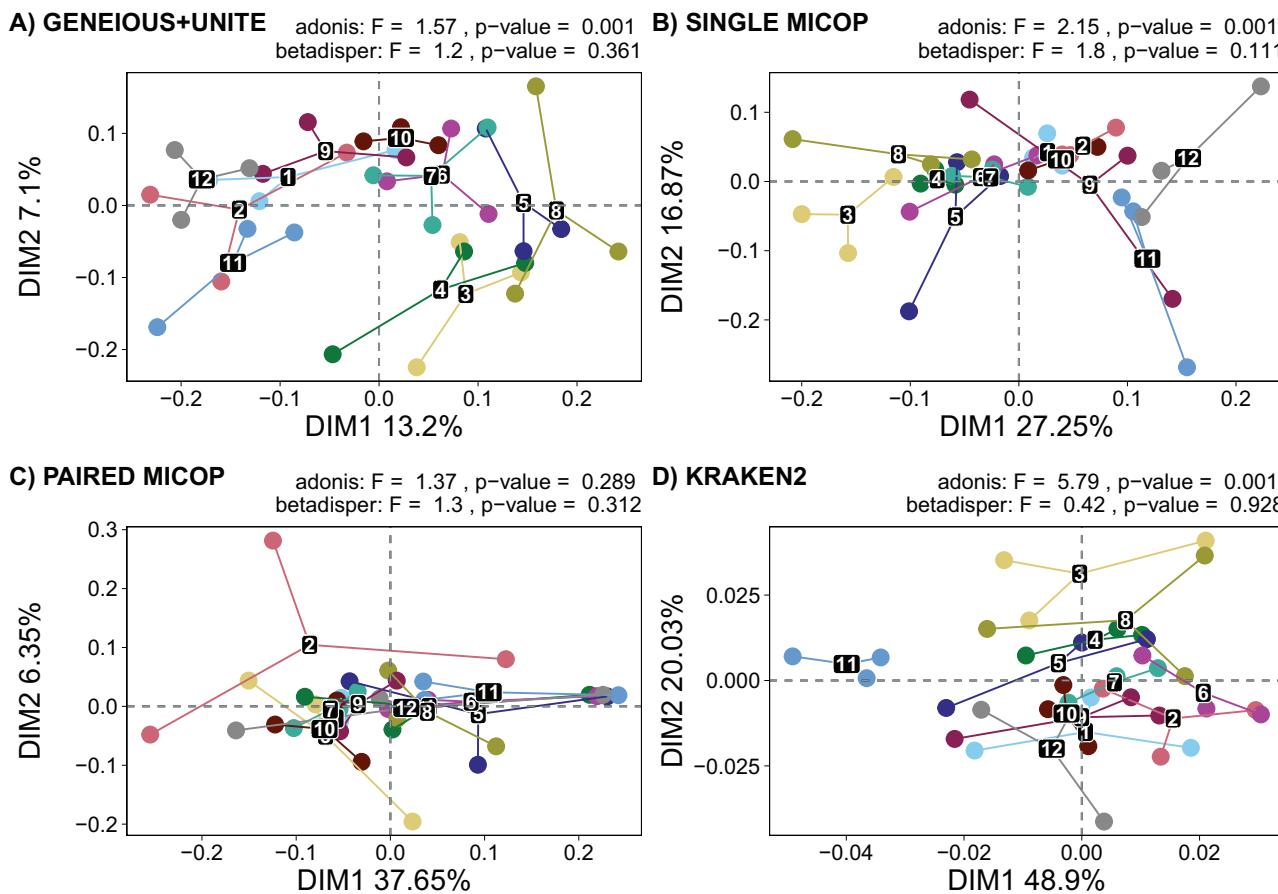
Full-size DOI: 10.7717/peerj.18323/fig-4

features), were similar across sites when using the Geneious+UNITE, Paired MiCoP, and Kraken2 methods, while Single MiCoP displayed distinct patterns across the sites.

### Taxonomic and functional assignment of fungi by sites affected by workflow

The taxonomic composition and abundance varied considerably with the workflow (Fig. 3). The most abundant genera differed across the different approaches, with only the genera *Aspergillus*, *Penicillium*, and *Trichoderma* detected by all workflows. For instance, Geneious+UNITE workflow detected *Russula* (mean relative abundance  $15.0 \pm 8.3\%$ ) and *Sebacina* ( $5.7 \pm 4.1\%$ ) as the most abundant genera, while the Single MiCoP *Penicillium* ( $10.1 \pm 4.7\%$ ) and *Phanerochaete* ( $5.7 \pm 2.0\%$ ) (Fig. 3). Paired MiCoP identified *Diplodia* as the dominant genus with a mean relative abundance of  $57.3 \pm 18.1\%$ , followed by *Enterocytozoon* ( $16.9 \pm 11.8\%$ ), while Kraken2 identified *Fusarium* ( $14 \pm 3.9\%$ ) as the most dominant genus followed by *Psilocybe* ( $7.6 \pm 2.9\%$ ).

Consistent with the analyses of taxonomic composition, functional assignments to guilds and trophic groups depended largely on the workflow used. The Geneious+UNITE workflow identified a greater number of taxonomic groups assigned to symbiotrophs, including ectomycorrhizal fungi genera, such as *Russula*, *Sebacina*, *Inocybe*, and



**Figure 5** Unconstrained principal coordinate analysis (PCoA) with Horn dissimilarities of the taxonomic structure of fungal communities (operational taxonomic units at 97% similarity or species) using different workflows: (A) GENEIOUS+UNITE, (B) SINGLE MICOP, (C) PAIRED MICOP and (D) KRAKEN2 (see Fig. 2). Variations in variance and homogeneity among sampling sites were determined with a permutational analysis of variance (adonis) and a permutational multivariate homogeneity of group dispersions analysis (betadisper). Numbers refer to the different sampling sites and their location is given in Fig. 1.

Full-size DOI: 10.7717/peerj.18323/fig-5

*Tomentella*, as well as some saprotrophs such as *Pleurotus* (Figs. 3 and 4). The Paired MiCoP approach exhibited a higher abundance of groups assigned to pathogens (e.g., *Clavispora/Candida* clade, *Enterocitiozoon*, and *Diplodia*). Single MiCoP detected a higher abundance of saprotrophs, especially wood saprotrophs (e.g., *Phanerochaete* and *Serpula*), while symbiotrophs and pathogens had a lower abundance. Kraken2 identified saprotrophs and pathotrophs, particularly plant pathogens, such as the pathogen/saprotroph *Fusarium* (Figs. 3 and 4).

### Dissimilarity patterns of soil fungal communities across sampling sites

Horn and Bray-Curtis metrics showed similar patterns (Figs. 5 and S2). The sampling sites clustered into three groups, *i.e.*, sites 1, 2, 9, 11, and 12, which are closer to the volcanoes of the National Parks *L. Malinche* and *Iztaccíhuatl-Popocatépetl*, sites 7 and 10 as a transition cluster, and sites 3, 4, 5, 6 and 8 which are in the transects between the National Parks.

These groups were more defined with Geneious+UNITE and Single MiCoP workflows and less with Kraken2 and Paired MiCoP workflows.

The analysis of dispersion showed that the variance was homogeneous across the sampling sites using both dissimilarity metrics ( $p > 0.05$ ). The perMANOVA test showed that the sampling site had a significant effect on the fungal taxonomic composition obtained by all workflows ( $p < 0.05$ ), except for Paired MiCoP with the Horn dissimilarity index (Fig. 5C).

### Fungal communities' similarity as affected by geographical distance

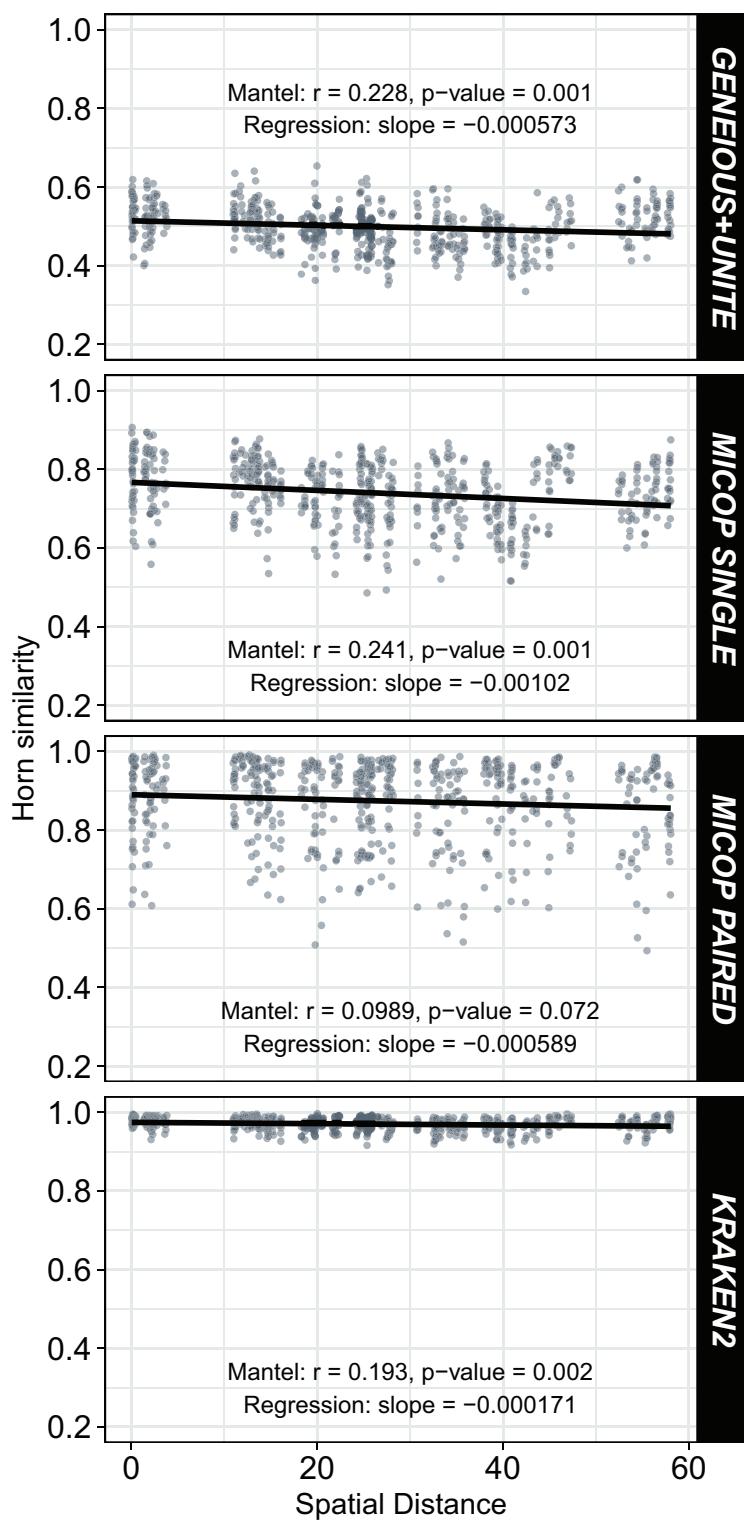
Weak but significant ( $p < 0.05$ ) negative correlations were observed between Horn similarities of fungal communities obtained by the different workflows and the spatial distance (Fig. 6). Absolute values of the Spearman Mantel correlations were in the order Single MiCoP ( $r = |-0.241|$ ) > Geneious+UNITE ( $r = |-0.228|$ ) > Kraken2 ( $r = |-0.193|$ ) > Paired MiCoP ( $r = |-0.0989|$ ). A different pattern was found with the Bray-Curtis similarities, although the negative correlations and significance were maintained: Geneious +UNITE ( $r = |-0.260|$ ) > Single MiCoP ( $r = |-0.213|$ ) > Kraken2 ( $r = |-0.212|$ ) > Paired MiCoP ( $r = |-0.177|$ ) (Fig. S3).

### Fungal communities' similarity as affected by environmental and vegetation characteristics

An exploratory PCA of environmental characteristics, *i.e.*, soil properties and nutrient content, showed two main clusters (Fig. S4). The first cluster comprised site samplings 1–2 and 9–12, located closer to the volcanoes in the National Parks *La Malinche* and *Iztaccíhuatl-Popocatépetl*. The second cluster consisted of sites 3–8, part of the transect between the National Parks. The first cluster was more spread out when compared to the second one, suggesting greater variations in environmental characteristics within the mountainous regions. Also, the first cluster was defined by higher P, N, and SOM content, and the second by higher WHC, pH, and content of clay, silt, Fe, K, and Cu.

Sites 1 and 12 were dominated by coniferous trees, sites 3–7 and 10–11 by broadleaf trees, while sites 2, 8–9 by conifers and broadleaves (Fig. S5A). Considering the tree genera, sites 1, 3–5 and 10–11 were dominated by one genus, the others by two or three (Fig. S5B). Accordingly, when considering vegetation and its characteristics, sampling sites were clustered into two groups, *i.e.*, sites 1–2 and 9–12 in one cluster and sites 3–8 in another (Fig. S6).

Correlations between the selected environmental and vegetation variables considered are given in Fig. S7. The CCA showed that the environmental variables that significantly ( $p < 0.05$ ) shaped the fungal communities differed across the bioinformatic workflow. Concentrations of Ca, loam, Mg, K, P and SOM, and pH and WHC significantly affected the soil fungal composition obtained by Geneious+UNITE workflow (Table 2). These variables, except the content of K, and P, shape the soil fungal composition obtained by Single MiCoP. For Paired MiCoP, the significant variables were SOM, Mg, P and N content, and pH, while for Kraken2 SOM, Ca and loam content, and pH. When accounting



**Figure 6** Relationship between Horn similarity of fungal communities obtained by different bioinformatic workflows (GENEIOUS+UNITE, SINGLE and PAIRED MICOP and KRAKEN2; see Fig. 2) and geographical distance. Mantel tests with Spearman's correlation were used to determine the association between community similarity and geographic matrices.

Full-size DOI: 10.7717/peerj.18323/fig-6

**Table 2** *p*-values of environmental variables resulted from ‘envfit’ using a canonical correlation analysis (CCA) on the Hellinger-transformed sequence data.

Environmental variable	Geneious+UNITE	Single MiCoP	Paired MiCoP	Kraken2
OM <sup>a</sup>	<b>0.015<sup>b</sup></b>	<b>0.004</b>	<b>0.029</b>	<b>0.001</b>
Ca	<b>0.001</b>	<b>0.001</b>	0.492	<b>0.001</b>
Silt	<b>0.001</b>	<b>0.001</b>	0.060	<b>0.033</b>
WHC <sup>c</sup>	<b>0.001</b>	<b>0.001</b>	0.105	0.378
Fe	0.064	0.481	0.118	0.451
K	<b>0.038</b>	0.080	0.786	0.786
Mg	<b>0.001</b>	<b>0.001</b>	<b>0.015</b>	0.084
pH	<b>0.001</b>	<b>0.001</b>	<b>0.001</b>	<b>0.001</b>
Cu	0.056	0.322	0.256	0.383
Clay	0.296	0.335	0.621	0.312
P	<b>0.021</b>	0.114	<b>0.051</b>	0.577
N	0.240	<b>0.001</b>	<b>0.012</b>	0.152

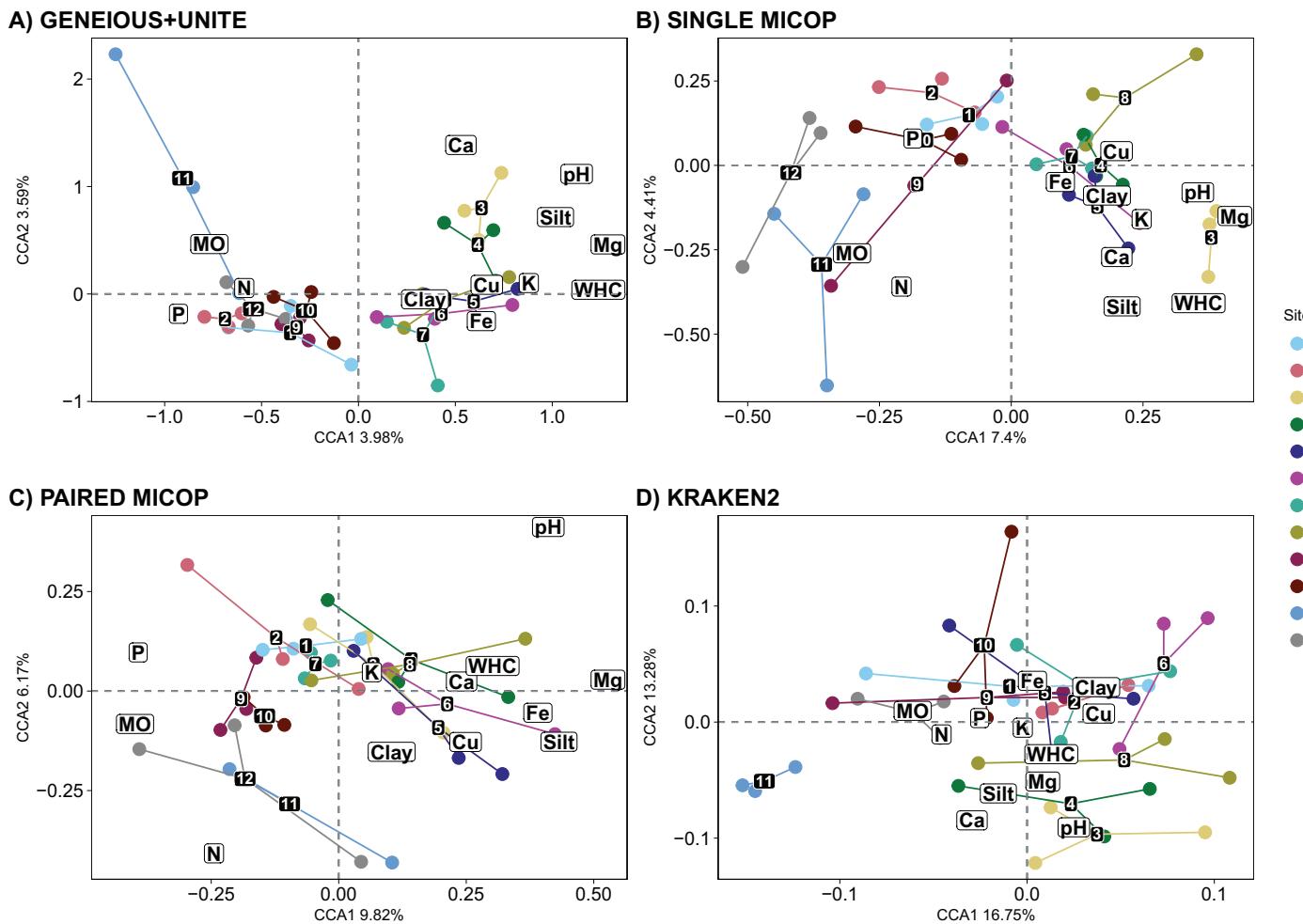
**Notes:**<sup>a</sup> OM, Organic material.<sup>b</sup> Values in bold are significant ( $p < 0.05$ ) and were retained for environmental distances analysis.<sup>c</sup> WHC, water holding capacity.

for the abundance of fungal groups (Horn, [Table S2](#), and Bray-Curtis, [Table S3](#)), Ca and SOM content, and WHC and pH, were significant when using Geneious+UNITE. For Single MiCoP the significant variables were WHC, pH and N content with the Horn distance matrix, and WHC, pH, and Mg and Ca content with Bray-Curtis.

The CCA plots with soil fungal groups as determined with the Geneious+UNITE and Single MiCoP workflow and the environmental variables gave more clearly separated clusters than those determined with Paired MiCoP and Kraken2 workflows ([Fig. 7](#)). These patterns were consistent, but less defined for Paired MiCoP and Kraken2 when accounting for abundance (Horn and Bray-Curtis distances) ([Figs. S8](#) and [S9](#)). The clustering pattern observed in this analysis was consistent with previous findings. Specifically, the sites located near the National Parks (1–2 and 9–12) formed one group, while the sites in the middle of the transect (3–8) constituted another group.

The significant vegetation traits also varied across workflows. The Geneious+UNITE workflow gave the highest number of significant traits with 11 in the CCA while Kraken2 had only 4 ([Table 3](#)). However, other results were obtained when accounting for abundance using Horn and Bray-Curtis distances ([Tables S4](#) and [S5](#)). Similar clustering patterns for the different sites were obtained across workflows for vegetation traits ([Figs. 8](#), [S10](#) and [S11](#)).

Partial Mantel tests exhibited higher and positive correlation values between fungal communities and the environmental distance matrix (ranging from 0.112 to 0.260 among the different workflows using Horn dissimilarity and from 0.164 to 0.280 using Bray-Curtis dissimilarity; [Tables S6](#) and [S7](#)) than with the vegetation distance matrix (ranging from -0.086 to 0.140 using Horn dissimilarity and from -0.098 to 0.178 using Bray-Curtis dissimilarity; [Tables S6](#) and [S7](#)).



**Figure 7** Canonical correspondence analysis (CCA) with the composition of soil fungal groups and environmental characteristics, *i.e.*, physicochemical soil properties and nutrient content, using different workflows: (A) GENEIOUS+UNITE, (B) SINGLE MICOP, (C) PAIRED MICOP and (D) KRAKEN2 (see Fig. 2). Numbers refer to the different sampling sites and their location is given in Fig. 1.

Full-size DOI: 10.7717/peerj.18323/fig-7

## DISCUSSION

In this study, we analyzed fungal communities in neotropical mountainous forest soils using environmental metagenomics. We aimed to identify the environmental, vegetation, and geographical factors controlling the spatial distribution of soil fungi on a regional scale. We used four different bioinformatic workflows for the fungal taxonomic profiling, and the results varied depending on the workflow used. This highlights the importance of a meticulous workflow selection and the use of suitable tools for an accurate analysis of fungal communities. It is important, however, to note that the fungal groups detected by all the workflows were characteristic of forest soils. Therefore, while misidentifications might be possible, the results were biased toward specific taxonomic or functional groups. For instance, when using the Geneious+UNITE workflow, we observed a high abundance of symbiotrophic fungi, such as the genera *Russula*, *Sebacina*, *Inocybe*, *Tomentella*, and *Cortinarius*. These are dominant and/or highly diverse ECM fungi in coniferous and

**Table 3** *p*-values of vegetation traits resulted from ‘envfit’ from canonical correlation analysis (CCA) on the Hellinger-transformed sequence data.

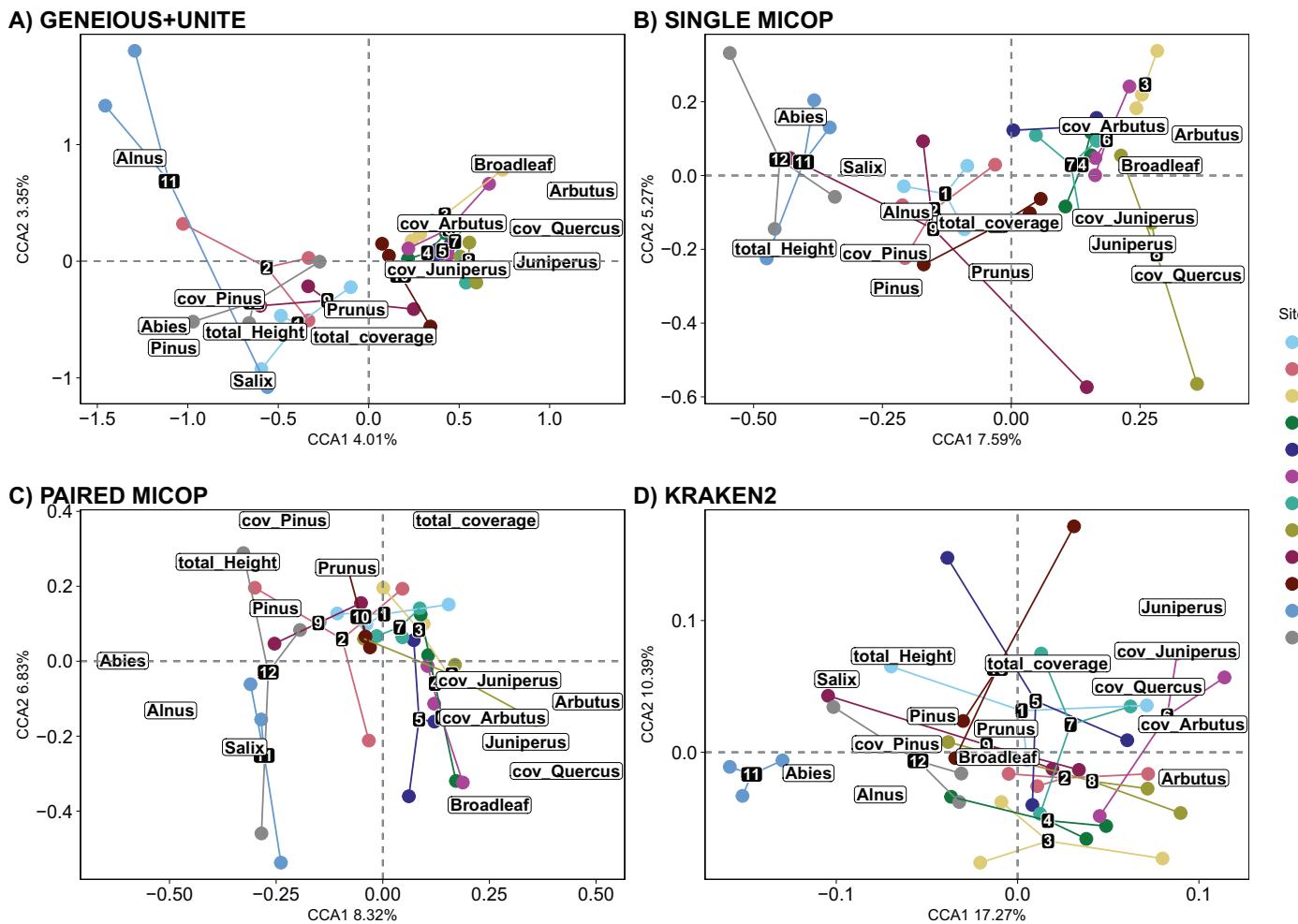
Vegetation variable	Geneious+ UNITE	Single MiCoP	Paired MiCoP	Kraken2
Proportion of broadleaf	<b>0.009<sup>a</sup></b>	<b>0.006</b>	0.090	0.878
Proportion of <i>Abies</i>	<b>0.001</b>	<b>0.001</b>	<b>0.001</b>	<b>0.018</b>
Proportion of <i>Pinus</i>	<b>0.001</b>	<b>0.001</b>	0.376	0.629
Proportion of <i>Salix</i>	<b>0.005</b>	0.054	<b>0.041</b>	<b>0.032</b>
Proportion of <i>Arbutus</i>	<b>0.001</b>	<b>0.001</b>	<b>0.012</b>	0.082
Proportion of <i>Alnus</i>	<b>0.001</b>	0.075	<b>0.014</b>	0.088
Proportion of <i>Juniperus</i>	<b>0.039</b>	<b>0.009</b>	0.052	<b>0.009</b>
Proportion of <i>Prunus</i>	0.584	0.126	0.482	0.941
Mean of total height	<b>0.003</b>	<b>0.001</b>	0.102	0.061
Mean of total coverage	0.130	0.118	0.179	0.168
Coverage of <i>Pinus</i>	<b>0.007</b>	<b>0.025</b>	<b>0.032</b>	0.361
Coverage of <i>Arbutus</i>	<b>0.018</b>	<b>0.035</b>	0.052	0.072
Coverage of <i>Quercus</i>	<b>0.002</b>	<b>0.001</b>	<b>0.007</b>	0.079
Coverage of <i>Juniperus</i>	0.217	0.133	0.169	<b>0.029</b>

**Note:**

<sup>a</sup> Values in bold are significant ( $p < 0.05$ ) and were retained for environmental distances analysis.

broadleaf forests from the TMVB (Reverchon *et al.*, 2012) and from other Nearctic (Barroetaveña, Cázares & Rajchenberg, 2007; Kennedy *et al.*, 2012) and Neotropical areas (Carranza *et al.*, 2018). They play a crucial role in forest ecosystems by forming mutualistic associations with tree roots, facilitating nutrient exchange, and enhancing plant growth and survival (Behie & Bidochka, 2014). The Single MiCoP approach revealed a higher abundance of saprotrophic fungi, particularly *Serpula*, commonly found on dead wood of various coniferous tree species (Carlsen *et al.*, 2011), and a lower abundance of EcM fungi, such as the genus *Tuber*, which has been reported to aid in host nutrient acquisition (Guevara *et al.*, 2013). In contrast, the Paired MiCoP and Kraken2 workflows detected pathogens including *Diplodia*, *Clavispore/Candida* clade, *Fusarium*, *Aspergillus*, and *Colletotrichum* at high frequency. *Diplodia* has been reported as a pine pathogen (Luchi *et al.*, 2014), and *Colletotrichum* is a well-known phytopathogen (da Silva *et al.*, 2020). *Fusarium* and *Aspergillus* are known plant pathogens (Syed, Tollamadugu & Lian, 2020). These variations in taxonomic and fungal group assignments may be attributed, at least in part, to the different types of databases used by each workflow, such as fungal genomes vs ITS genetic markers. For example, fewer and sometimes different taxonomic groups were detected when we used a genome-based database in the Kraken2 workflow (manually constructed with fungal taxa; Table S8).

Alpha diversity results varied greatly among workflows, and patterns were not consistent. This can be explained because each workflow represents the results of different taxonomic and functional groups, which do not necessarily follow the same patterns. The only consistent result was the lower diversity in the 11 and 12 sampling points (sites closer to the *Iztaccíhuatl* volcano). Similarly, no significant correlations were found between the



**Figure 8** Canonical correspondence analysis (CCA) of the composition of soil fungal groups and vegetation traits, *i.e.*, proportion of genera, total coverage and coverage by genus, and total height of the trees, using different workflows: (A) GENEIOUS+UNITE, (B) SINGLE MICOP, (C) PAIRED MICOP and (D) KRAKEN2 (see Fig. 2). Numbers refer to the different sampling sites and their location is given in Fig. 1.

Full-size DOI: 10.7717/peerj.18323/fig-8

alpha diversity obtained with the different workflows and the factors analyzed in this study (spatial distance, vegetation, and edaphic properties; data not shown).

Since the results of the different workflows detected different functional and taxonomic groups, we used these data to test hypotheses related to different functional groups of fungi. We found similar clustering patterns for the symbiophils (Geneious+UNITE), pathogens (Paired MiCoP), and saprotrophs (Single MiCoP), suggesting that environmental, vegetation, and dispersal limitation affected the composition of these fungal guilds to a similar extent. Clustering patterns when considering saprotrophs +pathogen fungi (Kraken2), however, were unclear.

Our study observed that sampling points closer to the National Parks clustered. This clustering can be attributed to the shared environmental and vegetation characteristics among these sites (Figs. S4 and S6). Although previous studies have identified geographical distance as an important predictor of soil fungal community composition globally

([Tedersoo et al., 2014](#)), spatial distance or dispersal limitation is not a major factor at the regional scale ([Xing et al., 2021](#)). Instead, the determinant factors at the regional scale are primarily related to environmental and vegetation/host characteristics ([Peay, Baraloto & Fine, 2013](#); [Tedersoo et al., 2020](#); [Xing et al., 2021](#)). However, it is worth noticing that our study revealed a decay in similarity with spatial distance, which can be attributed to the geological relief of the mountains ([Bahram et al., 2013](#)). Additionally, landscape heterogeneity resulting from land-use changes, such as those affecting the study areas, can decrease gene flow by creating geographic barriers or a matrix of unsuitable habitats that restrict dispersal ([Bisbing et al., 2021](#)). This suggests that dispersal limitation may still contribute to shaping fungal community composition and distribution, albeit not being the primary factor in our study.

The analysis of environmental characteristics showed that sites 1, 2, 9, 10, 11, and 12, closer to the National Parks, were associated with vital characteristics for forest environments, such as soil SOM, P, and N content. Intermediate sites 3, 4, 5, 6, 7, and 8 were associated with pH, Mg, Ca, Mg, Fe, WHC, Clay, Silt, and Cu. Intermediate sites were located in forest fragments surrounded by agricultural land. This spatial arrangement provides a plausible explanation for the observed patterns. Previous studies have reported a decline in N and SOM content when land use changes, *i.e.*, from forest to agriculture ([Campos et al., 2007](#); [Steenwerth et al., 2002](#)). Additionally, higher values for pH, and Ca and Mg concentrations have been reported in arable soil. This is attributed to the application of fertilizers, pesticides, and chemicals, such as calcium carbonate ( $\text{CaCO}_3$ ), contributing to the specific edaphic characteristics of these sites ([Pouyat et al., 2007](#); [Steenwerth et al., 2002](#)).

The pH, and SOM, loam, Ca, and Mg content were identified as common drivers of the fungal groups across all datasets, *i.e.*, symbiotrophs (Geneious+UNITE), saprotrophs (Single MiCoP), pathogens (Paired MiCoP) and saprotrophs+pathogens (Kraken2). For saprotrophs and pathogens, *i.e.*, as determined with Single MiCoP, Paired MiCoP, and Kraken2, total N emerged as an important driver. Although N content has been reported as an important driver shaping EcM taxa ([Yang et al., 2022](#)), we did not find this effect for symbiotrophs (Geneious+UNITE) in this study. The results obtained from Single MiCoP and Paired MiCoP workflows revealed SOM as the defining characteristic shaping saprotrophs. They specialize in decomposing organic matter, which makes it an important factor in their distribution and abundance.

The significance of pH as a driver of fungal communities on a regional scale agrees with previous reports ([Tedersoo et al., 2020](#); [Vasco-Palacios et al., 2020](#)). pH is a fundamental factor in the assembly and distribution of soil microorganisms across various spatial scales, from local to global ([Aslani et al., 2022](#); [Ge et al., 2017](#); [Genevieve et al., 2019](#); [Glassman, Wang & Bruns, 2017](#); [Liu et al., 2018](#)). It controls the capacity of soils to store or release nutrients as it affects the solubility of P, Fe, Mn, Al, and Ca. Soil pH is closely associated with soil Ca content and tree species' abundance, which affect fungal richness and composition ([Tedersoo et al., 2020](#)). Other factors, such as SOM or N content, emerge as important drivers of fungal communities, but their significance may vary based on the

location of the studied area. It would be interesting to investigate whether Ca consistently affects fungi across different scales and locations.

The number of vegetation traits that accounted for the taxonomic composition varied across the different datasets. For symbiotrophs (Geneious+UNITE), 12 vegetation traits were found to be significant, while nine were identified for saprotrophs (Single MiCoP), seven for pathogens (Paired MiCoP) and four for saprotrophs+pathogens (Kraken2). These findings align with previous studies indicating that vegetation plays a more prominent role in shaping the community structure of biotrophic fungi, such as symbiotrophs, compared to that of saprotrophic or pathogenic fungi (Janowski & Leski, 2022; Tedersoo *et al.*, 2020). In contrast, vegetation composition and configuration had a smaller effect on saprotrophs and pathogens. If the pathogenic fungi we identified had remained pathogens of plants instead of animals, then vegetation would have significantly influenced this fungal guild. Comparing the patterns identified in this study concerning saprotrophic fungi with other investigations is challenging, primarily because most studies tend to prioritize the analysis of symbiotrophic fungi, with less focus on saprotrophic fungi.

Notably, sites near both National Parks were characterized by coniferous vegetation, while broadleaf trees dominated intermediate sites. Previous studies at the regional scale have highlighted the importance of the relationship between fungal communities and vegetation in maintaining biological diversity (Yang *et al.*, 2017). Our findings support the significance of this association and propose that aboveground plant characteristics could serve as a valuable proxy for regional conservation planning. Therefore, considering vegetation characteristics is crucial when studying the distribution and diversity of fungal communities.

Interestingly, the genera *Russula*, *Sebacina*, *Inocybe*, *Tomentella* and *Cortinarius* are often dominant across the Nearctic, Palearctic and Neotropic zones (Barroeta & Vélez, 2007; Corrales *et al.*, 2022; Erlandson *et al.*, 2016; Li *et al.*, 2018; Reverchon *et al.*, 2012; Tedersoo *et al.*, 2022). Despite variations in their species assemblage, these fungi display a consistent global presence in diversity, prevalence, and dominance. The high abundance of *Diplodia* in the studied soil is a cause for concern, as it is a pathogen affecting trees and crops (Luchi *et al.*, 2014). Our findings suggest that the soil may serve as a reservoir for propagules of this important phytopathogen. *Enterocytozoon* is the most common microsporidia infection. Microsporidia, vertebrate intestinal commensals, are single-celled parasites with a broad host range and can become opportunistic pathogens. While not previously detected in soil metagenomic analyses (Donovan *et al.*, 2018), our study reveals its high abundance, which is unexpected given its life cycle. However, it might have originated from vertebrate feces as they were detected as propagules and not in their active form.

The National Parks used in this study are in areas affected by the megalopolis comprising Mexico City and its surrounding areas, whose rapid growth has significantly impacted the natural ecosystems. The extensive fragmentation and disturbance of natural vegetation have led to a potential loss of biodiversity (Tobón *et al.*, 2016). It is, therefore, crucial to understand the diversity and distribution of species in this region and identify

areas of high species richness (hotspots) and conservation areas to preserve its biodiversity. Fungal communities play a vital role in forest ecosystems, so it is essential to identify the factors that shape their composition and distribution, considering the potential negative impact of anthropogenic activities on them.

Metabarcoding approaches have been employed to obtain more information from environmental samples. These methods have limitations, such as time-consuming protocols and biases introduced by specific primers used. In contrast, shotgun metagenome sequencing is increasingly used. This method analyzes total eDNA, allowing for studying fungi, bacteria, viruses, and other organisms and their interactions, potential functions, and community structure. However, metagenomics is more costly, bioinformatically complex, and requires deep sequencing to reliably represent less abundant organisms, such as fungi. Several algorithms and programs to analyze metagenomic data have been developed in recent years, including BLAST, USEARCH, DIAMOND, and Kraken ([Alloui et al., 2015](#); [Buchfink, Xie & Huson, 2015](#); [Camacho et al., 2009](#); [Wood, Lu & Langmead, 2019](#)), which identify the most similar sequence of their database and trace it to a lower common ancestor (LCA). Additionally, new approaches such as MiCoP utilize fast-read mapping to construct a comprehensive reference database of full eukaryotic genomes, maximizing read usage ([LaPierre et al., 2019](#)). Other tools, such as QIIME2 ([Bolyen et al., 2019](#)), can be used if reads are first mapped against a reference database with programs like Geneious ([Kearse et al., 2012](#)). This approach has been used in species of lichens that are not frequently found in reference databases ([Smith et al., 2020](#)). Still, a significant challenge is the limited availability of pipelines or workflows for studying fungal communities compared to those for viral and bacterial communities ([LaPierre et al., 2019](#); [Taş et al., 2021](#)).

Bioinformatic workflows for taxonomic annotations of metagenomic data have fundamental differences. Some workflows like Kraken2 and MiCoP use fungal genome reference databases. This approach allows to identify almost all fungal metagenomic reads as fungal. Still, a drawback is that the limited number of available fungal genomes means that only a small number of taxa can be detected. On the other hand, other methods use the ITS marker gene database (*e.g.*, UNITE+Geneious), which can identify a larger number of taxa due to the extensive ITS sequence data available. However, it is limited to detecting only ITS regions in the metagenomes. Our results reflect these differences: MiCoP and Kraken2 identified more reads but fewer taxa, whereas UNITE+Geneious identified more taxa but fewer reads. Due to these fundamental differences among the methods, we can assume that results obtained will be different.

## CONCLUSIONS

We used four different bioinformatic workflows to profile fungal taxonomy and found variations in the results depending on the workflow used. Each workflow emphasized different characteristic fungal groups, *i.e.*, symbiotic, saprotrophic and pathogenic fungi. Environmental, vegetation and geographical factors affected the spatial distribution of soil fungi on a regional scale. pH, and SOM, Ca, Mg and loam content were common drivers of fungal communities independent of the workflow used. Vegetation traits shaped

the composition of symbiotic fungi more than that of saprotrophic and pathogenic fungi. This highlights the importance of considering vegetation traits when studying fungal community diversity and distribution. Clustering patterns of sampling points near National Parks, indicating shared environmental and vegetation traits. Dispersal limitation was found to contribute to fungal community composition, although it was not the primary factor in the study. Overall, this study provides important insights into the challenges and opportunities of studying fungal communities at a regional scale using metagenomic data. However, the workflow should be reported as it might be biased towards specific fungal guilds.

## ACKNOWLEDGEMENTS

To all ejidos, landowners, managers, and authorities who allowed us to conduct this study, Special thanks to ejido San Rafael Ixtapalucan, ejido San José Nanacamilpa, ejido Moxolahuac, ejido Hueyotlipan, Ejido San Gabriel and ejido La Caridad Cuaxonacayo. To the owners and managers of Tandi-Chala Adventure Park and Camping, Piedra Canteada Ecotourism Park and Paraje el Madroño. To the communities of Álvaro Obregón, San Francisco Mitepec, Tizatlán, el Carmen las Carrozas, La Ascensión Huitzcolotepec, San Ambrosio Texantla, San Francisco Tetlanohcan, San Rafael Tepatlaxco, San Esteban Tizatlán, Santa Cruz Tenancingo and San Felipe Hidalgo. To the authorities of the 16 municipalities (belonging to the states of Tlaxcala, Puebla and the State of Mexico): Amaxac de Guerrero, Chiautempan, Espanita, Hueyotlipan, Ixtacuixtla de Mariano Matamoros, Ixtapaluca, Nanacamilpa de Mariano Arista, Panotla, San Francisco Tetlanohcan, Tlahuapan, Teolocholco, Tlaxcala, Totolac and Xaltocan. To the authorities of “Comisión Nacional de Áreas Naturales Protegidas” (CONANP) for allowing us to work in the National Parks, to Secretaría de Medio Ambiente de Tlaxcala. To the project technicians M. en G. Lenin Ríos Figueroa and M. en C. María Chanel Juárez Ramírez that helped us in the field, with logistics and administrative matters. To Judith Galicia for her support with the administration of the project grant. The authors thank “Estación Científica La Malinche” for their logistic support during fieldwork; Ph.D. Ricardo Vazquez Juarez, “Laboratorio de Genómica y bioinformática” of CIBNOR and “Microbioma Lab” for his help in the bioinformatics analysis and access to computing resources.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This research was funded by “Fondo Sectorial de Investigación de la Secretaría de Relaciones Exteriores y el Consejo Nacional de Humanidades Ciencias y Tecnologías” (SRE-CONAHCYT) (Grant CAR 286794). Stephanie E Hereira-Pacheco received grant-aided support from CONAHCYT (grant number: 929602). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Grant Disclosures

The following grant information was disclosed by the authors:

Fondo Sectorial de Investigación de la Secretaría de Relaciones Exteriores y el Consejo Nacional de Humanidades Ciencias y Tecnologías (SRE-CONAHCYT): CAR 286794. CONAHCYT: 929602.

## Competing Interests

The authors declare that they have no competing interests.

## Author Contributions

- Stephanie Hereira-Pacheco performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Itzel Arias-Del Razo conceived and designed the experiments, authored or reviewed drafts of the article, funding acquisition and project administration, and approved the final draft.
- Alejandra Miranda-Carrazco analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Luc Dendooven analyzed the data, authored or reviewed drafts of the article, funding acquisition, and approved the final draft.
- Arturo Estrada-Torres conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Yendi E. Navarro-Noya conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

## DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The raw sequences are available at SRA NCBI: [PRJNA875082](https://www.ncbi.nlm.nih.gov/sra/PRJNA875082).

## Data Availability

The following information was supplied regarding data availability:

All scripts and metadata for statistical analyses and visualizations are available at Zenodo: Stephanie Elizabeth Hereira Pacheco. (2024). Steph0522/Fungal\_Metagenomes\_PNIP\_PNML: Latest version final (Final). Zenodo. <https://doi.org/10.5281/zenodo.13647231>.

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.18323#supplemental-information>.

## REFERENCES

Alloui T, Boussebough I, Chaoui A, Nouar AZ, Chettah MC. 2015. Usearch: a meta search engine based on a new result merging strategy. In: *2015 7th International Joint Conference on*

*Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*. Vol. 1531–536.

**Arriaga-Jiménez A, Rös M, Halffter G.** 2018. High variability of dung beetle diversity patterns at four mountains of the Trans-Mexican Volcanic Belt. *PeerJ* **6**(1):e4468 DOI [10.7717/peerj.4468](https://doi.org/10.7717/peerj.4468).

**Aslani F, Geisen S, Ning D, Tedersoo L, Bahram M.** 2022. Towards revealing the global diversity and community assembly of soil eukaryotes. *Ecology Letters* **25**(1):65–76 DOI [10.1111/ele.13904](https://doi.org/10.1111/ele.13904).

**Bahram M, Köljalg U, Courty P-E, Diédhiou AG, Kjøller R, Pölme S, Ryberg M, Veldre V, Tedersoo L.** 2013. The distance decay of similarity in communities of ectomycorrhizal fungi in different ecosystems and scales. In N. Wurzburger (Ed.). *Journal of Ecology* **101**(5):1335–1344 John Wiley & Sons, Ltd DOI [10.1111/1365-2745.12120](https://doi.org/10.1111/1365-2745.12120).

**Bahram M, Netherway T, Hildebrand F, Pritsch K, Drenkhan R, Loit K, Anslan S, Bork P, Tedersoo L.** 2020. Plant nutrient-acquisition strategies drive topsoil microbiome structure and function. *New Phytologist* **227**(4):1189–1199 DOI [10.1111/nph.16598](https://doi.org/10.1111/nph.16598).

**Barroetaveña C, Cázares E, Rajchenberg M.** 2007. Ectomycorrhizal fungi associated with ponderosa pine and Douglas-fir: a comparison of species richness in native western North American forests and Patagonian plantations from Argentina. *Mycorrhiza* **17**(5):355–373 DOI [10.1007/s00572-007-0121-x](https://doi.org/10.1007/s00572-007-0121-x).

**Behie SW, Bidochka MJ.** 2014. Nutrient transfer in plant-fungal symbioses. *Trends in Plant Science* **19**(11):734–740 DOI [10.1016/j.tplants.2014.06.007](https://doi.org/10.1016/j.tplants.2014.06.007).

**Bennett JW.** 1998. Mycotechnology: the role of fungi in biotechnology1 Based on a lecture held at the symposium, 'Progress in US Biotechnology', at the 8th European Congress on Biotechnology (ECB8) in Budapest, Hungary, August 1997.1. *Journal of Biotechnology* **66**(2):101–107 DOI [10.1016/S0168-1656\(98\)00133-3](https://doi.org/10.1016/S0168-1656(98)00133-3).

**Bhantana P, Rana MS, Sun X, Moussa MG, Saleem MH, Syaifudin M, Shah A, Poudel A, Pun AB, Bhat MA, Mandal DL, Shah S, Zhihao D, Tan Q, Hu C-X.** 2021. Arbuscular mycorrhizal fungi and its major role in plant growth, zinc nutrition, phosphorous regulation and phytoremediation. *Symbiosis* **84**(1):19–37 DOI [10.1007/s13199-021-00756-6](https://doi.org/10.1007/s13199-021-00756-6).

**Bisbing SM, Urza AK, Buma BJ, Cooper DJ, Matocq M, Angert AL.** 2021. Can long-lived species keep pace with climate change? Evidence of local persistence potential in a widespread conifer. *Diversity and Distributions* **27**(2):296–312 DOI [10.1111/ddi.13191](https://doi.org/10.1111/ddi.13191).

**Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15):2114–2120 DOI [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170).

**Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch AK, Jiang L, Kaehler B, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolka T, Kreps J, Langille MG, Lee J, Ley R, Liu Y, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton J, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS II, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CH, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG.** 2019. Reproducible, interactive,

scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37(8):852–857 DOI 10.1038/s41587-019-0209-9.

**Bouffaud M-L, Bernaud E, Colombet A, van Tuinen D, Wipf D, Redecker D. 2016.** Regional-scale analysis of arbuscular mycorrhizal fungi: the case of Burgundy vineyards. *OENO One* 50(1):1–8 DOI 10.20870/oeno-one.2016.50.1.49.

**Bremner JM. 1996.** Nitrogen-total. In: *Methods of Soil Analysis Part 3—Chemical Methods*. Hoboken: Wiley.

**Buchfink B, Xie C, Huson DH. 2015.** Fast and sensitive protein alignment using DIAMOND. *Nature Methods* 12(1):59–60 DOI 10.1038/nmeth.3176.

**Bushnell B. 2014.** BBMap: a fast, accurate, splice-aware aligner. Berkeley: Lawrence Berkeley National Lab. Available at <https://www.osti.gov/biblio/1241166>.

**Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.** BLAST+: architecture and applications. *BMC Bioinformatics* 10(1):421 DOI 10.1186/1471-2105-10-421.

**Campos CA, Oleschko LK, Etchevers BJ, Hidalgo MC. 2007.** Exploring the effect of changes in land use on soil quality on the eastern slope of the Cofre de Perote Volcano (Mexico). *Forest Ecology and Management* 248(3):174–182 DOI 10.1016/j.foreco.2007.05.004.

**Carlsen T, Engh IB, Decock C, Rajchenberg M, Kauserud H. 2011.** Multiple cryptic species with divergent substrate affinities in the *Serpula himantoides* species complex. *Fungal Biology* 115(1):54–61 DOI 10.1016/j.funbio.2010.10.004.

**Carranza VJ, DiStefano GJ, Walter-Marín M, Mata HM. 2018.** Comparative study of Macrofungi on dead oak wood in two neotropical montane forests in Costa Rica. *Polibotánica* 2018(45):35–56 DOI 10.18387/polibotanica.45.4.

**Cassel DK, Nielsen DR. 1986.** Field capacity and available water capacity. In: Klute A, ed. *Methods of Soil Analysis. Part I. Physical and Mineralogical Methods*, Agronomy Monograph No. 9. Madison: Soil Science Society of America, 901–926.

**Chao A, Chiu C-H, Jost L. 2014.** Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annual Review of Ecology, Evolution, and Systematics* 45(1):297–324 DOI 10.1146/annurev-ecolsys-120213-091540.

**Chaudhary VB, Nolimal S, Sosa-Hernández MA, Egan C, Kastens J. 2020.** Trait-based aerial dispersal of arbuscular mycorrhizal fungi. *New Phytologist* 228(1):238–252 DOI 10.1111/NPH.16667.

**CONABIO. 2023.** Suelo y vegetación. Available at <http://www.conabio.gob.mx>.

**Corrales A, Koch RA, Vasco-Palacios AM, Smith ME, Ge ZW, Henkel TW. 2022.** Diversity and distribution of tropical ectomycorrhizal fungi. *Mycologia* 114(6):919–933 DOI 10.1080/00275514.2022.2115284.

**da Silva LL, Moreno HLA, Correia HLN, Santana MF, de Queiroz MV. 2020.** *Colletotrichum*: species complexes, lifestyle, and peculiarities of some sources of genetic variability. *Applied Microbiology and Biotechnology* 104(5):1891–1904 DOI 10.1007/s00253-020-10363-y.

**Donovan PD, Gonzalez G, Higgins DG, Butler G, Ito K. 2018.** Identification of fungi in shotgun metagenomics datasets. *PLOS ONE* 13(2):e0192898 DOI 10.1371/JOURNAL.PONE.0192898.

**Erlandson SR, Savage JA, Cavender-Bares JM, Peay KG. 2016.** Soil moisture and chemistry influence diversity of ectomycorrhizal fungal communities associating with willow along an hydrologic gradient. *FEMS Microbiology Ecology* 92(1):148 DOI 10.1093/FEMSEC/FIV148.

**Gao C, Montoya L, Xu L, Madera M, Hollingsworth J, Purdom E, Singan V, Vogel J, Hutmacher RB, Dahlberg JA, Coleman-Derr D, Lemaux PG, Taylor JW.** 2020. Fungal community assembly in drought-stressed sorghum shows stochasticity, selection, and universal ecological dynamics. *Nature Communications* **11**(1):34 DOI [10.1038/s41467-019-13913-9](https://doi.org/10.1038/s41467-019-13913-9).

**Gee GW, Bauder JW.** 1986. Particle size analysis. *Methods of Soil Analysis: Part 1—Physical and Mineralogical Methods* **9**(9):901–926 DOI [10.2136/sssabookser5.1.2ed.c36](https://doi.org/10.2136/sssabookser5.1.2ed.c36).

**Ge ZW, Brenneman T, Bonito G, Smith ME.** 2017. Soil pH and mineral nutrients strongly influence truffles and other ectomycorrhizal fungi associated with commercial pecans (*Carya illinoiensis*). *Plant and Soil* **418**(1):493–505 DOI [10.1007/s11104-017-3312-z](https://doi.org/10.1007/s11104-017-3312-z).

**Genevieve L, Roxanne G, Amélie M, Danny B, Vincent M, Hugo G.** 2019. Estimation of fungal diversity and identification of major abiotic drivers influencing fungal richness and communities in Northern temperate and boreal Quebec forests. *Forests* **10**(12):1096 DOI [10.3390/f10121096](https://doi.org/10.3390/f10121096).

**Glassman SI, Wang IJ, Bruns TD.** 2017. Environmental filtering by pH and soil nutrients drives community assembly in fungi at fine spatial scales. *Molecular Ecology* **26**(24):6960–6973 DOI [10.1111/mec.14414](https://doi.org/10.1111/mec.14414).

**Gu Z, Eils R, Schlesner M.** 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**(18):2847–2849 DOI [10.1093/bioinformatics/btw313](https://doi.org/10.1093/bioinformatics/btw313).

**Guevara G, Bonito G, Trappe JM, Cázares E, Williams G, Healy RA, Schadt C, Vilgalys R.** 2013. New North American truffles (*Tuber* spp.) and their ectomycorrhizal associations. *Mycologia* **105**(1):194–209 DOI [10.3852/12-087](https://doi.org/10.3852/12-087).

**Hazard C, Gosling P, van der Gast CJ, Mitchell DT, Doohan FM, Bending GD.** 2013. The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. *The ISME Journal* **7**(3):498–508 DOI [10.1038/ismej.2012.127](https://doi.org/10.1038/ismej.2012.127).

**Hijmans RJ.** 2021. Package ‘geosphere’. Spherical trigonometry (1.5-14). Available at [CRAN.R-project.org/package=geosphere](https://CRAN.R-project.org/package=geosphere).

**Hui N, Liu X, Kotze DJ, Jumpponen A, Francini G, Setälä H.** 2017. Ectomycorrhizal fungal communities in urban parks are similar to those in natural forests but shaped by vegetation and park age. *Applied and Environmental Microbiology* **83**(23):109 DOI [10.1128/AEM.01797-17](https://doi.org/10.1128/AEM.01797-17).

**INEGI.** 2024. Climatología. Available at <https://www.inegi.org.mx/temas/climatologia/>.

**Janowski D, Leski T.** 2022. Factors in the distribution of mycorrhizal and soil fungi. *Diversity* **14**(12):1122 DOI [10.3390/d14121122](https://doi.org/10.3390/d14121122).

**Kahle D, Wickham H.** 2013. ggmap: spatial visualization with ggplot2. *The R Journal* **5**(1):144–161 DOI [10.32614/RJ-2013-014](https://doi.org/10.32614/RJ-2013-014).

**Kassambara A.** 2020. Package ‘ggpubr’. R package version 0.4.0 (Vol. 6). Available at <https://cran.r-project.org/web/packages/ggpubr/index.html>.

**Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12):1647–1649 DOI [10.1093/bioinformatics/bts199](https://doi.org/10.1093/bioinformatics/bts199).

**Kennedy PG, Smith DP, Horton TR, Molina RJ.** 2012. *Arbutus menziesii* (Ericaceae) facilitates regeneration dynamics in mixed evergreen forests by promoting mycorrhizal fungal diversity and host connectivity. *American Journal of Botany* **99**(10):1691–1701 DOI [10.3732/ajb.1200277](https://doi.org/10.3732/ajb.1200277).

**Kernaghan G, Widden P, Bergeron Y, Légaré S, Paré D.** 2003. Biotic and abiotic factors affecting ectomycorrhizal diversity in boreal mixed-woods. *Oikos* **102**(3):497–504 DOI [10.1034/j.1600-0706.2003.12415.x](https://doi.org/10.1034/j.1600-0706.2003.12415.x).

**Khalid M, Ur Rahman S, Liu X, Rehman A, Jumpponen A, Johan Kotze D, Setälä H, Hui N.** 2024. Exploring the impact of urbanization and vegetation type on fungal communities: insights into divergent, mycorrhizal, and saprophytic associations driven by climate patterns. *CATENA* 238(2):107860 DOI [10.1016/j.catena.2024.107860](https://doi.org/10.1016/j.catena.2024.107860).

**Kivlin SN, Winston GC, Goulden ML, Treseder KK.** 2014. Environmental filtering affects soil fungal community composition more than dispersal limitation at regional scales. *Fungal Ecology* 12(C):14–25 DOI [10.1016/j.funeco.2014.04.004](https://doi.org/10.1016/j.funeco.2014.04.004).

**Lan G, Quan F, Yang C, Sun R, Chen B, Zhang X, Wu Z.** 2022. Driving factors for soil fungal and bacterial community assembly in tropical forest of China. *Applied Soil Ecology* 177(1):104520 DOI [10.1016/j.apsoil.2022.104520](https://doi.org/10.1016/j.apsoil.2022.104520).

**LaPierre N, Mangul S, Alser M, Mandric I, Wu NC, Koslicki D, Eskin E.** 2019. MiCoP: microbial community profiling method for detecting viral and fungal organisms in metagenomic samples. *BMC Genomics* 20(5):423 DOI [10.1186/s12864-019-5699-9](https://doi.org/10.1186/s12864-019-5699-9).

**Li H.** 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv preprint DOI [10.48550/arXiv.1303.3997](https://doi.org/10.48550/arXiv.1303.3997).

**Li D.** 2018. hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. *Journal of Open Source Software* 3(31):1041 DOI [10.21105/joss.01041](https://doi.org/10.21105/joss.01041).

**Li J, Chen L, Wang H, Ouyang S, Liu X, Lu J.** 2022. Pattern and drivers of soil fungal community along elevation gradient in the *Abies georgei* forests of Segila mountains, Southeast Tibet. *Global Ecology and Conservation* 39:e02291 DOI [10.1016/j.gecco.2022.e02291](https://doi.org/10.1016/j.gecco.2022.e02291).

**Li H, Guo J, Karunarathna SC, Ye L, Xu J, Hyde KD, Mortimer PE.** 2018. Native forests have a higher diversity of macrofungi than comparable plantation forests in the Greater Mekong Subregion. *Forests* 9(7):402 DOI [10.3390/F9070402](https://doi.org/10.3390/F9070402).

**Liu D, Liu G, Chen L, Wang J, Zhang L.** 2018. Soil pH determines fungal diversity along an elevation gradient in Southwestern China. *Science China Life Sciences* 61(6):718–726 DOI [10.1007/s11427-017-9200-1](https://doi.org/10.1007/s11427-017-9200-1).

**Lomolino MV, Riddle BR, Brown JH.** 2006. *Biogeography*. Third Edition. Sunderland, MA: Sinauer Associates, Inc.

**Lu J, Breitwieser FP, Thielen P, Salzberg SL.** 2017. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science* 3:e104 DOI [10.7717/peerj-cs.104](https://doi.org/10.7717/peerj-cs.104).

**Luchi N, Oliveira Longa CM, Danti R, Capretti P, Maresi G.** 2014. *Diplodia sapinea*: the main fungal species involved in the colonization of pine shoots in Italy. *Forest Pathology* 44(5):372–381 DOI [10.1111/efp.12109](https://doi.org/10.1111/efp.12109).

**Mastretta-Yanes A, Xue AT, Moreno-Letelier A, Jorgensen TH, Alvarez N, Piñero D, Emerson BC.** 2018. Long-term in situ persistence of biodiversity in tropical sky islands revealed by landscape genomics. *Molecular Ecology* 27(2):432–448 DOI [10.1111/mec.14461](https://doi.org/10.1111/mec.14461).

**McMurdie PJ, Holmes S.** 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE* 8(4):e61217 DOI [10.1371/journal.pone.0061217](https://doi.org/10.1371/journal.pone.0061217).

**Moctezuma V, Halffter G, Escobar F.** 2016. Response of copronecrophagous beetle communities to habitat disturbance in two mountains of the Mexican Transition Zone: influence of historical and ecological factors. *Journal of Insect Conservation* 20(6):945–956 DOI [10.1007/s10841-016-9923-5](https://doi.org/10.1007/s10841-016-9923-5).

**Newbold T, Oppenheimer P, Etard A, Williams JJ.** 2020. Tropical and Mediterranean biodiversity is disproportionately sensitive to land-use and climate change. *Nature Ecology & Evolution* 4(12):1630–1638 DOI [10.1038/s41559-020-01303-0](https://doi.org/10.1038/s41559-020-01303-0).

**Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, Schilling JS, Kennedy PG. 2016.** FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology* 20:241–248 DOI [10.1016/j.funeco.2015.06.006](https://doi.org/10.1016/j.funeco.2015.06.006).

**Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Köljalg U, Abarenkov K. 2019.** The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research* 47(D1):D259–D264 DOI [10.1093/nar/gky1022](https://doi.org/10.1093/nar/gky1022).

**Nowosad J. 2023.** Package ‘rcartocolor’. R package version 2.1.1. Available at <https://CRAN.R-project.org/package=rcartocolor>.

**Oksanen FJ, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Solymos P, Stevens MHH, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Antoniazi Evangelista HB, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill MO, Lahti L, McGlinn D, Ouellette M-H, Ribeiro Cunha E, Smith T, Stier A, Ter Braak CJF, Weedon J. 2017.** Vegan: Community Ecology Package. R package Version 2.4-3. Available at <https://CRAN.R-project.org/package=vegan>.

**Paradis E, Claude J, Strimmer K. 2004.** APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20(2):289–290 DOI [10.1093/bioinformatics/btg412](https://doi.org/10.1093/bioinformatics/btg412).

**Peay KG, Baraloto C, Fine PVA. 2013.** Strong coupling of plant and fungal community structure across western Amazonian rainforests. *The ISME Journal* 7(9):1852–1861 DOI [10.1038/ismej.2013.66](https://doi.org/10.1038/ismej.2013.66).

**Peay KG, Bruns TD. 2014.** Spore dispersal of basidiomycete fungi at the landscape scale is driven by stochastic and deterministic processes and generates variability in plant-fungal interactions. *New Phytologist* 204(1):180–191 DOI [10.1111/nph.12906](https://doi.org/10.1111/nph.12906).

**Pouyat RV, Yesilonis ID, Russell-Anelli J, Neerchal NK. 2007.** Soil chemical and physical properties that differentiate urban land-use and cover types. *Soil Science Society of America Journal* 71(3):1010–1019 DOI [10.2136/SSAJ2006.0164](https://doi.org/10.2136/SSAJ2006.0164).

**R Core Team. 2023.** R: a language and environment for statistical computing. Available at <https://www.r-project.org/>.

**Rahbek C, Gotelli NJ, Colwell RK, Entsminger GL, Rangel TFLVB, Graves GR. 2006.** Predicting continental-scale patterns of bird species richness with spatially explicit models. *Proceedings of the Royal Society B: Biological Sciences* 274(1607):165–174 DOI [10.1098/rspb.2006.3700](https://doi.org/10.1098/rspb.2006.3700).

**Rasmussen PU, Hugerth LW, Blanchet FG, Andersson AF, Lindahl BD, Tack AJM. 2018.** Multiscale patterns and drivers of arbuscular mycorrhizal fungal communities in the roots and root-associated soil of a wild perennial herb. *New Phytologist* 220(4):1248–1261 DOI [10.1111/nph.15088](https://doi.org/10.1111/nph.15088).

**Reverchon F, del Pilar Ortega-Larrocea M, Bonilla-Rosso G, Pérez-Moreno J. 2012.** Structure and species composition of ectomycorrhizal fungal communities colonizing seedlings and adult trees of *Pinus montezumae* in Mexican neotropical forests. *FEMS Microbiology Ecology* 80(2):479–487 DOI [10.1111/j.1574-6941.2012.01314.x](https://doi.org/10.1111/j.1574-6941.2012.01314.x).

**Rodríguez P, Ochoa-Ochoa LM, Munguía M, Sánchez-Cordero V, Navarro-Sigüenza AG, Flores-Villela OA, Nakamura M. 2019.** Environmental heterogeneity explains coarse-scale β-diversity of terrestrial vertebrates in Mexico. *PLOS ONE* 14(1):e0210890 DOI [10.1371/journal.pone.0210890](https://doi.org/10.1371/journal.pone.0210890).

**Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016.** VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4(17):e2584 DOI [10.7717/peerj.2584](https://doi.org/10.7717/peerj.2584).

**Ruiz-Sanchez E, Specht CD.** 2013. Influence of the geological history of the Trans-Mexican Volcanic Belt on the diversification of *Nolina parviflora* (Asparagaceae: Nolinoideae). *Journal of Biogeography* **40**(7):1336–1347 DOI [10.1111/jbi.12073](https://doi.org/10.1111/jbi.12073).

**Salinas N, Cosio EG, Silman M, Meir P, Nottingham AT, Roman-Cuesta RM, Malhi Y.** 2021. Editorial: tropical montane forests in a changing environment. *Frontiers in Plant Science* **12**:712748 DOI [10.3389/fpls.2021.712748](https://doi.org/10.3389/fpls.2021.712748).

**Smith HB, Dal Grande F, Muggia L, Keuler R, Divakar PK, Grewe F, Schmitt I, Lumbsch HT, Leavitt SD.** 2020. Metagenomic data reveal diverse fungal and algal communities associated with the lichen symbiosis. *Symbiosis* **82**(1):133–147 DOI [10.1007/s13199-020-00699-4](https://doi.org/10.1007/s13199-020-00699-4).

**Steenwerth KL, Jackson LE, Calderón FJ, Stromberg MR, Scow KM.** 2002. Soil microbial community composition and land use history in cultivated and grassland ecosystems of coastal California. *Soil Biology and Biochemistry* **34**(11):1599–1611 DOI [10.1016/S0038-0717\(02\)00144-X](https://doi.org/10.1016/S0038-0717(02)00144-X).

**Sui X, Zeng X, Li M, Weng X, Frey B, Yang L, Li M.** 2022. Influence of different vegetation types on soil physicochemical parameters and fungal communities. *Microorganisms* **10**(4):829 DOI [10.3390/microorganisms10040829](https://doi.org/10.3390/microorganisms10040829).

**Syed S, Tollamadugu NVKVP, Lian B.** 2020. *Aspergillus* and *Fusarium* control in the early stages of *Arachis hypogaea* (groundnut crop) by plant growth-promoting rhizobacteria (PGPR) consortium. *Microbiological Research* **240**(1):126562 DOI [10.1016/j.micres.2020.126562](https://doi.org/10.1016/j.micres.2020.126562).

**Taş N, de Jong AEE, Li Y, Trubl G, Xue Y, Dove NC.** 2021. Metagenomic tools in microbial ecology research. *Current Opinion in Biotechnology* **67**:184–191 DOI [10.1016/j.copbio.2021.01.019](https://doi.org/10.1016/j.copbio.2021.01.019).

**Tedersoo L, Anslan S, Bahram M, Drenkhan R, Pritsch K, Buegger F, Padari A, Mikryukov V, Gohar D, Amiri R, Hiiesalu I, Lutter R, Rosenvald R, Rähn E, Adamson K, Drenkhan T, Tullus H, Jürimaa K, Sibul I, Hagh-Doust N, Otsing E, Põlme S, Metslaid M, Loit K, Agan A, Puusepp R, Varik I, Kõlalg U, 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**:561190 DOI [10.3389/fmicb.2020.01953](https://doi.org/10.3389/fmicb.2020.01953).

**Tedersoo L, Bahram M, Põlme S, Kõlalg U, Yorou NS, Wijesundera R, Villarreal Ruiz L, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Rosas M, Riit T, Ratkowsky D, Pritsch K, Pöldmaa K, Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Majuakim L, Lodge DJ, Lee SS, Larsson KH, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo LD, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, De Kesel A, Dang T, Chen X, Buegger F, Brearley FQ, Bonito G, Anslan S, Abell S, Abarenkov K.** 2014. Global diversity and geography of soil fungi. *Nature* **346**(6213):1256688 DOI [10.1126/science.1256688](https://doi.org/10.1126/science.1256688).

**Tedersoo L, Bahram M, Zinger L, Nilsson RH, Kennedy PG, Yang T, Anslan S, Mikryukov V.** 2022. Best practices in metabarcoding of fungi: from experimental design to results. *Molecular Ecology* **31**(10):2769–2795 DOI [10.1111/mec.16460](https://doi.org/10.1111/mec.16460).

**Tedersoo L, May TW, Smith ME.** 2010. Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. *Mycorrhiza* **20**(4):217–263 DOI [10.1007/s00572-009-0274-x](https://doi.org/10.1007/s00572-009-0274-x).

**Thomas GW, Sparks DL, Page AL, Helmke PA, Loepert RH, Soltanpour PN, Tabatabai MA, Johnston CT, Sumner ME.** 1996. Soil pH and soil acidity. In: *Methods of Soil Analysis. Part 3- Chemical Methods*. Hoboken: Wiley, 475–490.

**Tobón W, Koleff P, Urquiza-Haas T, Méndez GG. 2016.** Propuesta metodológica para identificar prioridades de restauración en México. In: *Experiencias Mexicanas en la Restauración de los Ecosistemas*. Cuernavaca: Universidad Nacional Autónoma de México, Centro Regional de Investigaciones, 31.

**Vasco-Palacios AM, Bahram M, Boekhout T, Tedersoo L. 2020.** Carbon content and pH as important drivers of fungal community structure in three Amazon forests. *Plant and Soil* 450(1):111–131 DOI [10.1007/s11104-019-04218-3](https://doi.org/10.1007/s11104-019-04218-3).

**Walkley A, Black IA. 1934.** An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* 37(1):29–38 DOI [10.1097/00010694-193401000-00003](https://doi.org/10.1097/00010694-193401000-00003).

**Ward EB, Duguid MC, Kuebbing SE, Lendemer JC, Bradford MA. 2022.** The functional role of ericoid mycorrhizal plants and fungi on carbon and nitrogen dynamics in forests. *New Phytologist* 235(5):1701–1718 DOI [10.1111/nph.18307](https://doi.org/10.1111/nph.18307).

**Wei T, Simko V. 2021.** R package corrplot: visualization of a correlation matrix} (0.92). Available at <https://github.com/taiyun/corrplot>.

**Wickham H. 2011.** ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* 3(2):180–185 DOI [10.1002/wics.147](https://doi.org/10.1002/wics.147).

**Wood DE, Lu J, Langmead B. 2019.** Improved metagenomic analysis with Kraken 2. *Genome Biology* 20(1):257 DOI [10.1186/s13059-019-1891-0](https://doi.org/10.1186/s13059-019-1891-0).

**Xing R, Gao Q, Zhang F, Wang J, Chen S. 2021.** Environmental filtering affects fungal communities more than dispersal limitation in a high-elevation hyperarid basin on Qinghai-Tibet Plateau. *FEMS Microbiology Letters* 368(6):fnab033 DOI [10.1093/femsle/fnab033](https://doi.org/10.1093/femsle/fnab033).

**Xu X, Chen C, Zhang Z, Sun Z, Chen Y, Jiang J, Shen Z. 2017.** The influence of environmental factors on communities of arbuscular mycorrhizal fungi associated with Chenopodium ambrosioides revealed by MiSeq sequencing investigation. *Scientific Reports* 7(1):45134 DOI [10.1038/srep45134](https://doi.org/10.1038/srep45134).

**Yang T, Adams JM, Shi Y, He J, Jing X, Chen L, Tedersoo L, Chu H. 2017.** Soil fungal diversity in natural grasslands of the Tibetan Plateau: associations with plant diversity and productivity. *New Phytologist* 215(2):756–765 DOI [10.1111/nph.14606](https://doi.org/10.1111/nph.14606).

**Yang N, Hua J, Zhang J, Liu D, Bhople P, Li X, Zhang Y, Ruan H, Xing W, Mao L. 2022.** Soil nutrients and plant diversity affect ectomycorrhizal fungal community structure and functional traits across three subalpine coniferous forests. *Frontiers in Microbiology* 13:1016610 DOI [10.3389/FMICB.2022.1016610/BIBTEX](https://doi.org/10.3389/FMICB.2022.1016610).

**Zhao J, Gao Q, Zhou J, Wang M, Liang Y, Sun B, Chu H, Yang Y. 2019.** The scale dependence of fungal community distribution in paddy soil driven by stochastic and deterministic processes. *Fungal Ecology* 42:100856 DOI [10.1016/j.funeco.2019.07.010](https://doi.org/10.1016/j.funeco.2019.07.010).