

RESEARCH

Open Access



# Metabarcoding with Illumina and Oxford Nanopore Technologies provides complementary insights into tree seed mycobiota

Jana Mittelstrass<sup>1\*</sup>, Renate Heinzelmann<sup>1</sup>, René Eschen<sup>2</sup>, Martin Hartmann<sup>3</sup>, Quirin Kupper<sup>1</sup>, Salome Schneider<sup>1</sup>, Simone Prospero<sup>1</sup> and Iva Franić<sup>1</sup>

## Abstract

**Background** Culturing of fungi is labor-intensive and reveals limited diversity, while high-throughput sequencing of barcodes (i.e., metabarcoding) enables a simultaneous detection of fungi from multiple environmental samples. Metabarcoding using short-read sequencers, such as Illumina platforms, provides high sequencing depths but results in many unidentified taxa. Long-read sequencing can improve species and genus assignments but might encompass lower sequencing depth and limit diversity coverage. In this study, fungi in seeds of eleven angiosperm and gymnosperm tree species were assessed using traditional culturing, Illumina short-read metabarcoding, and Oxford Nanopore Technologies long-read metabarcoding. We focused on seed-borne fungi as understanding their diversity and potential impacts on seedlings is crucial for securing plant health. We compared (1) the number and identity of fungal genera and species between metabarcoding approaches and traditional culturing and (2) fungal alpha- and beta-diversity between metabarcoding methods, considering different hosts and fungal lifestyles.

**Results** In both short- and long-read metabarcoding datasets, similar numbers of fungal reads and operational taxonomic units were assigned to comparable numbers of fungal genera and species. About one-third of the identified genera were plant pathogens, followed by saprotrophs and endophytes. Culturing overall revealed fewer fungal genera, while most of the fungal reads in short-read metabarcoding datasets stemmed from cultured taxa. Long-read metabarcoding revealed lower per-sample diversity than short-read metabarcoding and distinct fungal communities compared to those from the short-read datasets. Host-dependent patterns in alpha- and beta-diversity were observed across methods, with angiosperms harboring more fungal taxa than gymnosperms, and distinct community structuring across host tree groups and species, although the differences were stronger in short-read than long-read metabarcoding datasets.

**Conclusions** Illumina and Oxford Nanopore Technologies metabarcoding captured similar host-dependent diversity patterns despite observed differences in numbers and composition of fungi. Short-read metabarcoding might be optimal for fungal biodiversity studies due to higher sequencing depths and resultant breadth of diversity. As error rates are continuing to decrease, reference databases expand, and throughput improves, long-read metabarcoding

\*Correspondence:

Jana Mittelstrass

jana.orbach@wsl.ch

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

is becoming a strong candidate for future diagnostic studies of fungi. Traditional culturing captures most of the fungi from short-read metabarcoding and remains valuable for obtaining isolates for further research.

**Keywords** High-throughput sequencing, Amplicon sequencing, Read length, MinION, ITS, Endophytes, Pathogens, Fungi, Seed mycobiome

## Background

Plant seeds harbor diverse fungal communities, some inherited from the mother plant and some acquired from the environment [1]. As a result, seed mycobiota (i.e., fungal spores and hyphae inhabiting seeds) exhibit host-driven diversity patterns, determined by host phylogeny and plant traits, and spatio-temporal diversity patterns driven by abiotic factors such as geographic location and climate [2–4]. Tree seed mycobiota, although historically understudied, have been the focus of an increasing number of microbiome studies in recent years. The studies explored the diversity of seed-borne fungi to assess the risks of pathogen movement with tree seed trade [5, 6], to understand the relative importance of the host and environment in shaping diversity of tree seed fungi [3, 7, 8] and to unravel the assembly processes in tree seedlings [9]. Growing interest in tree seed mycobiota is mainly due to the increasingly recognized implications seed-associated organisms have for ecosystem functioning and services, and for commercial forestry [10, 11]. For example, some important tree pathogens are known to be seed-borne [10]. The pathogen *Fusarium circinatum*, the causal agent of pine pitch canker, and *Diplodia sapinea*, an opportunistic fungal pathogen causing Diplodia tip blight, have been linked to the movement of pine seeds with devastating consequences for pine plantations [12–14]. However, not all fungi associated with seeds are detrimental to seed and seedling health and development. In fact, there is growing evidence that seed and seedling development may benefit from some seed fungi [1, 15, 16] through improved seed quality, plant growth promotion and seedling development [17, 18], or increased resistance to biotic and abiotic stresses. Thus, the development of efficient tools to detect and identify seed mycobiota is critical to better understand the potential of these complex communities to reduce or improve plant performance.

Since the second half of the nineteenth century, pro- and eukaryotic microorganisms, including fungi associated with plants, have been studied using culturing methods. In these methods, specific fungi are isolated from their original substrate, grown in pure culture on nutrient media, observed under the microscope, described, and sometimes used for further experiments (e.g., to test interactions with the host or to determine the host range). Using culturing methods, fungi are identified

either based on morphological characteristics of their cultures, fruiting bodies, or spores [19, 20], and—since the second half of the twentieth century—by DNA-based methods. The identification of cultures through Sanger sequencing has been widely employed for decades [21]. However, obligate biotrophs, such as powdery mildews, downy mildews and rusts depend on living host cells and thus cannot be cultured [22]. Similarly, isolating rare or slow-growing fungi [21, 23] can be done but requires meticulous culturing work and patience. Some of the limitations associated with culturing techniques can be addressed by using high-throughput sequencing (HTS) technologies.

High-throughput sequencing of molecular marker genes (i.e., metabarcoding) allows a rapid parallel assessment of many individuals, including obligate biotrophs and rare taxa, directly from host tissues. Technological advances [24, 25], population of reference sequence databases [26] and the reduction of sequencing costs [27] have ensured a broad user base for platforms such as Illumina MiSeq, resulting in a myriad of studies revealing unprecedented microbial diversity in various biotic [8, 28–32] and abiotic environments [33, 34]. At the same time, the key challenges associated with metabarcoding include the need for adequate laboratory and analytical infrastructure, particularly bioinformatic analysis [35], and incomplete reference sequence databases for taxonomic identification [36–38]. To ensure amplification of the target taxa, special care must be taken in the selection of appropriate barcoding markers [39, 40], while, in case of plant-associated microbiota, reducing host interference [43].

A major bottleneck of metabarcoding on the Illumina sequencing platform is that it yields short amplicons of up to maximum ~550 bases in length using paired-end approaches. The fungal marker of choice, the internal transcribed spacer (ITS) region, spans up to 700 base pairs (bp) in most fungi [41, 42] but shows considerable interspecific size variation of multiple hundred bp [43, 44]. Current Illumina technology only allows the sequencing of one of the ITS subregions (i.e., ITS1 or ITS2). Albeit both ITS subregions are sufficiently variable for a broad range of fungi [41], depending on the primers or ITS region used, taxonomic biases may arise [42]. Thus, targeting the entire ITS at once may provide greater taxonomic resolution [45] and improve species

identification. Technological advances are now fuelling the further development of novel long-read sequencing technologies, such as Pacific Biosciences [46] or Oxford Nanopore Technologies (ONT) [47]. ONT sequencing using the MinION platform can be done in-house or even in field settings [48], at economic prices and can deliver reads of tens to hundreds of kilobases [45, 49]. Long sequences can increase chances of correct species identification [50] by capturing more phylogenetic information in longer amplicons, which is also critical for accurate assignment of fungal lifestyles [51]. These novel long-read sequencing technologies have been methodologically validated using bacterial [52] and fungal [36, 53, 54] mock communities with strains expected in the studied environments, and ONT has been applied for the characterization of water- [55] or plant-associated fungi [47, 50, 56], among others. In line with other studies that have compared the influence of methodology on assessed mycobiomes, a comparative study of the ONT and Illumina platform, alongside traditional culturing methods, would provide a comprehensive understanding of differently assessed fungal communities when done across tree species or origins.

The main objective of this study was to systematically compare seed mycobiota of eleven angio- and gymnosperm tree species determined by (i) short-read Illumina metabarcoding using two different primer pairs targeting ITS2, (ii) long-read ONT metabarcoding targeting the full-length ITS region and (iii) traditional fungal culturing with Sanger sequencing, to assess whether different fungal assessment approaches capture fungi with similar taxon and lifestyle assignments, and to determine whether host-dependent diversity patterns are preserved across methods. We hypothesized that long-read

sequencing would reveal lower per-sample diversity due to lower sequencing depth but a higher taxonomic resolution resulting in more identified fungal genera and species than short-read sequencing (ONT vs. Illumina). Moreover, we expected culturing to reveal a smaller fraction of the total community than the metabarcoding approaches.

Methods

Sample acquisition

The analyzed seed samples were obtained in 2016 [6] and belonged to 58 commercially traded seed lots of five angiosperm and six gymnosperm tree species originating from three continents (Table 1). Fungi were assessed from a total of 100 seeds per seed lot. Seeds were surface-sterilized with sodium hypochlorite (0.5%) for 5 min, subsequently rinsed twice with sterile water and air-dried under sterile conditions [6]. Each seed was cut in half and one half was used for fungal culturing (see below), while the other halves were pooled per seed lot and subjected to DNA extraction using the DNeasy PowerPlant Pro Kit (Qiagen, Hilden, Germany). The DNA was used as a template for Illumina MiSeq sequencing of two different ITS2 amplicons (‘Taylor’ and ‘Tedersoo’) and for ONT MinION sequencing, spanning the whole ITS region (Supplementary Table S1). While the ‘Taylor’ dataset and the culturing data were previously described in Franić et al. [6], the ‘Tedersoo’ and ‘ONT’ datasets were examined together in this study.

Fungal assessment by metabarcoding

Illumina MiSeq sequencing

Two Illumina ITS2 libraries were generated and bioinformatically processed using protocols and methodology

**Table 1** Numbers of seed lots analyzed using each method and their distributions across host tree species, host tree group and their continent of origin

| Tree species               | Tree group   | Origin        | Illumina meta-<br>barcoding | ONT meta-barcoding | Culturing     |
|----------------------------|--------------|---------------|-----------------------------|--------------------|---------------|
| <i>Acer macrophyllum</i>   | Angiosperm   | North America | 6                           | 4                  | 5             |
| <i>Acer palmatum</i>       | Angiosperm   | Asia          | 5                           | 5                  | 5             |
| <i>Acer pseudoplatanus</i> | Angiosperm   | Europe        | 3                           | 2                  | 3             |
| <i>Fagus sylvatica</i>     | Angiosperm   | Europe        | 6                           | 5                  | 5             |
| <i>Quercus garryana</i>    | Angiosperm   | North America | 6                           | 5                  | 6             |
| <i>Larix gmelinii</i>      | Gymnosperm   | Asia          | 3                           | 3                  | 3             |
| <i>Picea abies</i>         | Gymnosperm   | Europe        | 5                           | 4                  | 4             |
| <i>Pinus ponderosa</i>     | Gymnosperm   | North America | 8                           | 7                  | 6             |
| <i>Pinus sylvestris</i>    | Gymnosperm   | Europe        | 5                           | 3                  | 4             |
| <i>Pinus tabulaeformis</i> | Gymnosperm   | Asia          | 4                           | 4                  | 4             |
| <i>Tsuga heterophylla</i>  | Gymnosperm   | North America | 7                           | 6                  | 5             |
| <i>n</i> = 11              | <i>n</i> = 2 | <i>n</i> = 3  | <i>n</i> = 58               | <i>n</i> = 48      | <i>n</i> = 50 |

previously published by Franić et al. [6]. For library preparation, the DNA of each sample (seed lot) was quantified using the Qubit dsDNA BR Assay Kit (Invitrogen, Thermo Fisher Scientific Inc., Waltham, MA, USA). After dilution to 10 ng/μL, 5 μL of template DNA were amplified in PCR with both primer sets (for primer overview and sequences, see Supplementary Table S1) in triplicates as described in Franić et al. [6]. Briefly, for the Taylor dataset, the primer pair *5.8S-Fun/ITS4-Fun* [57] was used in PCR reactions with a total volume of 20 μL using the JumpStart REDTaq ReadyMix Reaction Mix (Sigma Aldrich, St. Louis, MO, USA). PCRs were run on a Veriti 96-Well Thermal Cycler (Applied Biosystems, Thermo Fisher Scientific Inc.) for 2 min at 94 °C; followed by 35 cycles of 30 s at 94 °C, 30 s at 58 °C, and 2 min at 72 °C; and a final extension of 10 min at 72 °C. For the Tedersoo library, PCR was carried out using the forward primer *ITS3ngs-mix* [33, 39] and the reverse primer *ITS4ngsUni* [40] with otherwise equal cycling conditions to those described above, except for an annealing temperature of 55 °C and a final concentration of 0.5 mM MgCl<sub>2</sub> in the reaction. Success of amplification, and lack of amplification in the negative (non-template) controls, was verified by agarose gel electrophoresis. The PCR products for both Taylor and Tedersoo amplicons were sent for library construction and sequencing to the Génome Québec Innovation Center (McGill University, Montréal, Canada) where all samples were sequenced on an Illumina MiSeq (Illumina Inc., San Diego, CA, USA) instrument using v3 chemistry (2×300 bp).

A customized bioinformatic pipeline, which is based on UPARSE [58] and described in detail in Franić et al. [6], was used for processing the demultiplexed files for both datasets. Briefly, the demultiplexed sequences were merged, then, PhiX and primer sequences were filtered out [59, 60]. Sequences were then quality-filtered and denoised, the ITS2 region was extracted using ITSx [61], and resulting sequences were clustered at 98% into unique Operational Taxonomic Units (OTUs) using UNOISE [62]. For the current study, these OTUs were taxonomically (re-)classified with a Naïve-Bayes classifier implemented in QIIME2 (version qiime2-2022.2 [63]) using the fungal UNITE database (version 10.05.21 [38]). To test the proportion of reads originating from the host trees, the UNITE database including other eukaryotic reference sequences ('All eukaryotes') was used. The raw sequences of the Tedersoo dataset were deposited at the European Nucleotide Archive (ENA) under accession number PRJEB81281, while raw sequences of the Taylor dataset were deposited in the Sequence Read Archive under accession number PRJNA550270 [6]. To minimize the presence of artefacts due to sequencing errors [49], singletons (i.e., OTUs with one read count) were

excluded from analyses, unless stated otherwise. In both Illumina metabarcoding datasets, singletons amounted to below 2.5% of the total reads (i.e., Taylor: 2.2%, Tedersoo: 0.63%).

#### ONT MinION sequencing

From 48 out of 58 DNA extracts produced in Franić et al. [6], libraries of the ribosomal RNA gene (rRNA) were prepared and sequenced on a MinION instrument (Oxford Nanopore Technologies, Oxford, UK), using a FLO-MIN106 flow cell (R9.4.1 pore chemistry). The primers *NSIshort/RCA95m* [36] were used for the amplification of a ~4–6 kB fragment, including the SSU, the full ITS region and part of the LSU of the rRNA gene (Supplementary Table S1). PCR reactions were run in three technical replicates. For each reaction, 5 μL of DNA template previously diluted to 10 ng/μL was added to the reaction mix. This consisted of 1×PrimeSTAR GXL Buffer (including 5 mM Mg<sup>2+</sup>), 0.8 mM dNTPs, 0.3 mg/mL BSA (Sigma Aldrich, Steinheim, Germany), 400 nM of each primer, and 0.5 U PrimeSTAR GXL DNA Polymerase (Takara Bio Europe, Saint-Germain-en-Laye, France), which was then filled up to the final reaction volume (20 μL) with PCR-grade water (Merck, Darmstadt, Germany). The following parameters were used to run PCR on a Veriti thermocycler (Applied Biosystems, Waltham, MA, USA): 1 min denaturing at 98 °C, followed by 29 cycles of 98 °C for 10 s, 55 °C for 15 s, 68 °C for 2.5 min. The pooled technical PCR replicates for each sample were visually inspected in gel electrophoresis. For samples that did not produce a visible amplicon band, DNA was diluted 1:20 v/v prior to PCR repetition. Each sample was then cleaned up using AMPure Beads XP (Beckman Coulter, Brea, CA, USA) following the manufacturer's instructions with small modifications (bead ratio of 0.5 (v/v), 80% ethanol, elution of the DNA after incubation at 37 °C for 15 min).

Unique 24-bp barcodes (Supplementary Table S2, EXP-PBC096 barcoding kit, ONT) were added in a limited-cycle PCR to each sample. The cycling conditions consisted of an initial denaturation step of 1 min at 98 °C, followed by 15 cycles of 98 °C for 10 s, 60 °C for 15 s, 68 °C for 4 min, and a final elongation step at 68 °C for 10 min. From each sample, 7 μL of the cleaned PCR product were used as template in a total reaction volume of 50 μL, including the same cycling conditions described above but using 1 mM dNTP Mix, 1.25U PrimeSTAR GXL DNA Polymerase and 200 nM of the barcoded primer. Resulting barcoded products were cleaned-up with AMPure Beads XP as described above and then pooled equimolarly: ~1 μg of the pooled libraries was used as input for the library preparation with the Ligation Sequencing Kit SQK-LSK109 (Oxford



Nanopore Technologies, Oxford, UK). Each sequencing run included 12 samples (Supplementary Table S2) and the flow cell used for runs 2 and 3 was washed between runs according to the manufacturer's instructions.

The software *guppy* (ONT, version 4.5.4) was used for the basecalling and demultiplexing of the raw reads. The numbers of reads per sample after demultiplexing (Supplementary Table S2) yielded a total of 7.98 M sequences. These sequences were further processed on high-performance clusters of the Federal Institute of Technology Zurich (ETH), Switzerland and the Swiss Federal Institute for Forest, Snow and Landscape Research WSL. The reads were re-oriented with *seqtk* (version 1.2-r94; <https://github.com/lh3/seqtk.git>), and the primers were detected and trimmed with *cutadapt* (version 1.12 [60]) with the error rate parameter *-e* set to 2. Only the reads in which both the forward and reverse primers were found in the correct orientation were retained. The trimmed reads were then filtered using *prinseq-lite* (version 0.20.4 [64]). Only sequences with lengths > 1 kB and < 8 kB, and GC contents between 40 and 55 were retained.

After these processing steps, the full-length ITS region was extracted from the filtered sequences using *ITSx* (version 1.1.3 [61]). The parameter *-E* was set to 0.001 and the complement-parameter was set to TRUE. We extracted the ITS region from the ONT sequences to be able to compare the sequences of all datasets among each other, although this meant discarding parts of the LSU and SSU included in the ONT sequence. Although these markers can contribute to the identification of fungi by phylogenetic placement, this approach is not yet commonly used and needs further development [65]. Then, the following steps of ONT data processing were chosen to mimic steps (i.e., clustering and mapping) undertaken in the Illumina data-processing pipeline (see Franić et al. [6]), but also considering the inferior sequence quality (i.e., choosing different parameters) of the ONT dataset due to high error rates in the ONT version used [66, 67]. The filtered, full-length ITS sequences were concatenated and further processed with *vsearch* v2.22.1 [68]. First, the sequences were pre-clustered with the *cluster\_size* algorithm using a cluster ID of 0.75 while restricting the minimum and maximum sequence length to 250 bp and 1 kB, respectively. These parameters for sequence length were chosen to minimize presence of host tree sequences. Using the *sortbysize* and *uchime\_denovo* algorithms, first global singleton sequences and subsequently chimeras were removed. After relabeling with *fastx\_filter* to include the prefix 'OTU\_', the initial sequences were matched to the OTUs with the *usearch\_global* algorithm, resulting in the final OTU table. During this step, the 'matching id' parameter was set to 0.97, while 'maxhits' was set to 1 and 'maxaccepts' to zero. This step was necessary for

the generation of the ONT OTU table, but also led to the loss of reads in the ONT dataset: in the ONT dataset, OTUs were almost exclusively singletons due to error rates higher than the clustering threshold, an issue that is known to be associated with the former ONT chemistry (versions on flow cells < R10.0) [66]. Consequently, further ONT analyses were based on all obtained reads per sample without excluding the singletons, as it was done for the Illumina datasets.

Lastly, as described for the Illumina metabarcoding datasets, the sequences of the ONT OTUs were classified against the fungal UNITE database (version 10.05.21 [38]) using a Naïve-Bayes classifier in QIIME2 (version qiime2-2022.2 [63]). The demultiplexed sequences are deposited at the ENA under the accession number PRJEB81455.

### Fungal assessment by culturing

From 50 out of the 58 seed lots analyzed with Illumina MiSeq sequencing (Table 1), fungi were also assessed by culturing. As described in more detail in Franić et al. [6], a total of 100 seed halves per seed lot were incubated at room temperature on 1.5% water agar (PPA, Pronadisa Lab Conda, Madrid, Spain) including 100 mg/L streptomycin to remove bacterial contaminants. All fungi emerging from seeds within 21 days were transferred onto streptomycin-containing (100 mg/L) potato dextrose agar (Difco Bacto PDA, 39 g/L; ChemieBrunschwig AG, Basel, Switzerland). All isolates obtained from seeds belonging to the same tree genus were grouped together based on their macromorphology. Two to five representative isolates per morphotype (in total 441 isolates) were then identified by Sanger sequencing of their ITS region as described by Franić et al. [6] using the primers *ITS1* and *ITS4* [69] (Supplementary Table S1). For this study, the edited ITS sequences of the morphotypes published in Franić et al. [6] and deposited under GenBank accession numbers MN105153–MN105593 were used. The sequences were re-classified against the same UNITE database used for the metabarcoding datasets (Illumina and ONT) as described above.

### Fungal lifestyle assignment

Lifestyles (i.e., the functional traits) of the fungal OTUs obtained with the Illumina MiSeq metabarcoding and ONT metabarcoding were determined for fungi identified at genus level using the interactive lookup table FungalTraits, following the author's instructions [70]. We note here that these functional assignments are based on the literature and reflect the most commonly occurring lifestyles for a given fungal taxon (i.e., genus) [70]. FungalTraits is an updated and extended version of FUNGuild [71], which was shown to perform better

in assigning fungal taxa to plant pathogens than FUN-Guild [72]. To discern plant endophytes (i.e., defined by Plôme et al. [70] as ‘asymptomatic, commensal or weakly mutualistic inhabitants’), plant pathogens and plant saprotrophs from other fungal lifestyle groups, we used the column ‘primary lifestyle’. Foliar and root endophytes were merged into the group ‘plant endophytes’, while litter, wood and nectar/tap saprotrophs were merged into the group ‘plant saprotrophs’. Entries of plant pathogens were defined as such.

### Data analysis

We first compared the overall number and taxonomic and lifestyle composition of fungal genera revealed by each of three metabarcoding datasets (Illumina MiSeq Taylor, Illumina MiSeq Tedersoo, ONT) and the culturing dataset. Afterwards, fungal alpha- and beta-diversity, considering OTUs, genus-level taxonomic and lifestyle assignments, were compared between metabarcoding approaches for different host tree groups and host tree species, with the latter representing a combination of host tree group and continent of origin, because seeds from single host tree species always originated from the same continent. The data analyses were conducted using *R* (version 4.2.0, [73]) in the programming environment *RStudio* [74]. To allow comparisons of composition and diversity between datasets, we focused on genus-level assignments across the Illumina, ONT and culturing datasets, as the lowest fungal taxonomic units in each method (i.e., OTUs, morphotypes) were generated separately and were thus not directly comparable.

### Fungal community comparison between metabarcoding and culturing datasets

The total number of shared and unique fungal genera was compared between culturing and the metabarcoding datasets. Moreover, the taxonomic composition of seed fungi revealed by the different metabarcoding methods was visualized, also focusing on genus-level assignments. Fungal OTUs obtained from each method were grouped by genus, after which relative abundances for each fungal genus in each metabarcoding dataset were calculated. Relative abundances of the twelve most abundant fungal genera were then plotted across host tree groups (i.e., angiosperms and gymnosperms). To test whether certain genera were more likely to occur in either host group and if this was persistent across methods, we conducted enrichment analyses for each metabarcoding dataset separately. For this, Illumina datasets were rarefied (resampled) to 1,000 reads per sample and the ONT dataset to 164 reads. To determine if the employed sequencing depth sufficiently captured OTU and genus richness, rarefaction curves were calculated and visualized for the

datasets (Supplementary Fig. S1). The rarefaction was done to account for differences in number of reads per sample within each dataset, and values were chosen to retain most of the samples (i.e., all samples in Illumina and around 62% of the samples in the ONT dataset). Quasi-Poisson generalized linear models (GLMs) were fit with the *R* function ‘glm’ to assess the differences in rarefied abundances of fungal genera (i.e., response variable) between angiosperms and gymnosperm (i.e., host tree group; explanatory variable) for each dataset separately. The use of GLMs for analyses of differential abundances in such way [75] is an alternative to the often-employed *R*-packages *Deseq2* [76], and can be done using custom scripts, for example, see Mittelstrass et al. [77].

For a direct comparison of fungi on a sequence-base, the representative sequences of the Taylor, Tedersoo and ONT OTUs were matched against the ITS sequences and corresponding taxonomic assignments of the fungal cultures with BLAST. For this, the classification algorithm *classify-consensus-blast* [78] implemented in QIIME2 was used, with the *perc-identity* parameter set to 0.85, and the ‘maxaccepts’ parameter set to 5. This analysis was used to show what relative proportion of overall sequences in each metabarcoding dataset was represented by the cultures, without an interference of taxonomic assignment bias through the database or classification algorithm.

Besides describing fungal communities across samples based on their dominant fungal genera, this study investigated relative abundances of genera assigned to different fungal lifestyles (i.e., plant endophytes, saprotrophs and pathogens) between samples belonging to different host tree groups and sequencing methods. As described for fungal genera, to test whether certain lifestyles were enriched in either host group, Illumina datasets were rarefied to 1,000 reads per sample and 164 reads in the ONT dataset, and quasi-Poisson GLMs were used to estimate the prevalence of the fungal lifestyles for each host group.

### Fungal alpha-diversity comparison between host tree groups and metabarcoding datasets

For the Illumina Taylor, Tedersoo and ONT metabarcoding datasets, alpha-diversity was measured as OTU and genus richness (i.e., the total number of OTUs or genera present in each sample). As richness is a metric that gives the same weight to abundant and rare taxa [79], it captures patterns of alpha-diversity of the whole fungal communities. In addition to this occurrence-based metric, alpha-diversity was estimated using abundance-weighted indices, i.e., the Shannon’s entropy and the Inverse Simpson Diversity [79, 80], allowing us to explore alpha-diversity patterns of dominant fungal communities. Abundance-weighted alpha-diversity metrics were

calculated using the R package *hillR* [81], and only for Illumina data. In the case of the ONT dataset, abundance-based measures could not be calculated because all OTUs in the ONT dataset were singletons.

Effects of the metabarcoding method (i.e., Taylor, Tedersoo and ONT), host tree group (i.e., angiosperms and gymnosperms) and their interaction on fungal alpha-diversity (i.e., OTU and genus richness) were tested in generalized linear mixed models (GLMMs) using the 'glmmTMB' function from the R package *glmmTMB* [82]. We included host tree species nested in host tree group in the models as a random factor. We assumed a zero-truncated negative binomial distribution for the errors of OTU richness, which were non-zero counts, and a negative binomial distribution for the errors of genus richness which contained one sample with no identified genera. Although correcting for differences in the number of reads among samples by setting the offset argument to the log of the number of sequencing reads is recommended for alpha-diversity models [77, 83], in the main text we only present the results of the models without offset correction. Due to dataset-inherent differences in the number of reads per sample (i.e., lower numbers of reads per sample in ONT vs. Illumina), comparing alpha-diversity between normalized datasets can result in an overestimation of diversity in ONT vs. Illumina datasets. We present results of the models run with the offset parameter for read normalization in the Supplementary materials (Additional file 2). Significance of factors in each model was tested using the 'Anova' function from the *car* package [84]. A Tukey comparison of least-squares means was performed with the function 'emmeans' from the *emmeans* package [85]. The same function was used to calculate the estimated marginal means and standard errors.

To test how the proportion of genera assigned to the different lifestyles was influenced by host tree group and metabarcoding method, we used Chi-Square tests on binomial GLMs [86, 87], across the full sample set and using the formula 'proportion ~ lifestyle + host group \* method' with the R function 'glm'. Proportions in this case were used instead of absolute numbers to account for differences in genus richness between datasets.

#### **Fungal beta-diversity comparison between host tree groups and metabarcoding datasets**

To visually explore the fungal beta-diversity within each dataset, Principal Component Analysis (PCA) and hierarchical clustering were used. For both analyses, the read count table was normalized to centered-log ratio (CLR [88]), a method that corrects for the sparse nature of compositional data. For PCA, the Monte-Carlo instances were generated and obtained using the functions 'aldex.

clr' and 'getMonteCarloInstances' from the R package *ALDEx2* [89]. The resulting normalized communities were then restricted to the 150 most abundant taxa in each sample for the analysis of beta-diversity at OTU level. As the 150 most abundant taxa partially varied between samples, this number added up to more than 150 taxa for each dataset. We also conducted a genus-level PCA, for which all genera identified in each dataset were used. The function 'prcomp' of the R *stats* package [73] was used to conduct PCA.

For community-level analyses using hierarchical clustering at genus level, each dataset was subset to the 50 most abundant genera across samples after CLR normalization. Heatmaps were then plotted with the R package *pheatmap* [90] using the Minkowski distance, a measure that describes the shortest distance between points in Euclidean space, as clustering method [91, 92].

The effects of 'host tree group' and 'method', and 'host tree species' and 'method', including interaction terms for both, on differences in fungal community structure (i.e., beta-diversity) at genus level were assessed with permutational multivariate analysis of variance (PERMANOVA) [93] using the 'adonis' function from the *vegan* package [94]. PERMANOVAs were run on the Aitchison dissimilarity matrix [88], which was calculated from merged raw abundance matrices containing OTUs assigned to genera across samples and datasets, using the 'vegdist' function from the *vegan* package with the method argument set to 'robust.aitchinson'. Focusing on genus assignments allowed us to analyze the metabarcoding datasets at once, as merging datasets at OTU level would not have been possible due to the separately generated and not fully overlapping amplicons represented by those OTUs.

Additionally, to determine pairwise differences between samples belonging to different host tree groups but that were assessed with different methods, we used the 'pairwise.adonis' function with the method argument set to 'fdr' for *P*-value adjustments from the *pairwiseAdonis* R package [95], which is a wrapper for multilevel pairwise comparison of the 'adonis' function from the R package *vegan*. We focus only on pairwise comparisons at host tree group and method level because the number of comparisons for host tree species and method would result in 529 pairwise comparisons.

## **Results**

### **Fungi revealed by the metabarcoding and culturing datasets**

After bioinformatic processing, Illumina MiSeq sequencing yielded a similar total number of reads for both datasets (Taylor: 2.49 M reads; Tedersoo: 2.48 M reads). Among those reads, which clustered into 1,391 and

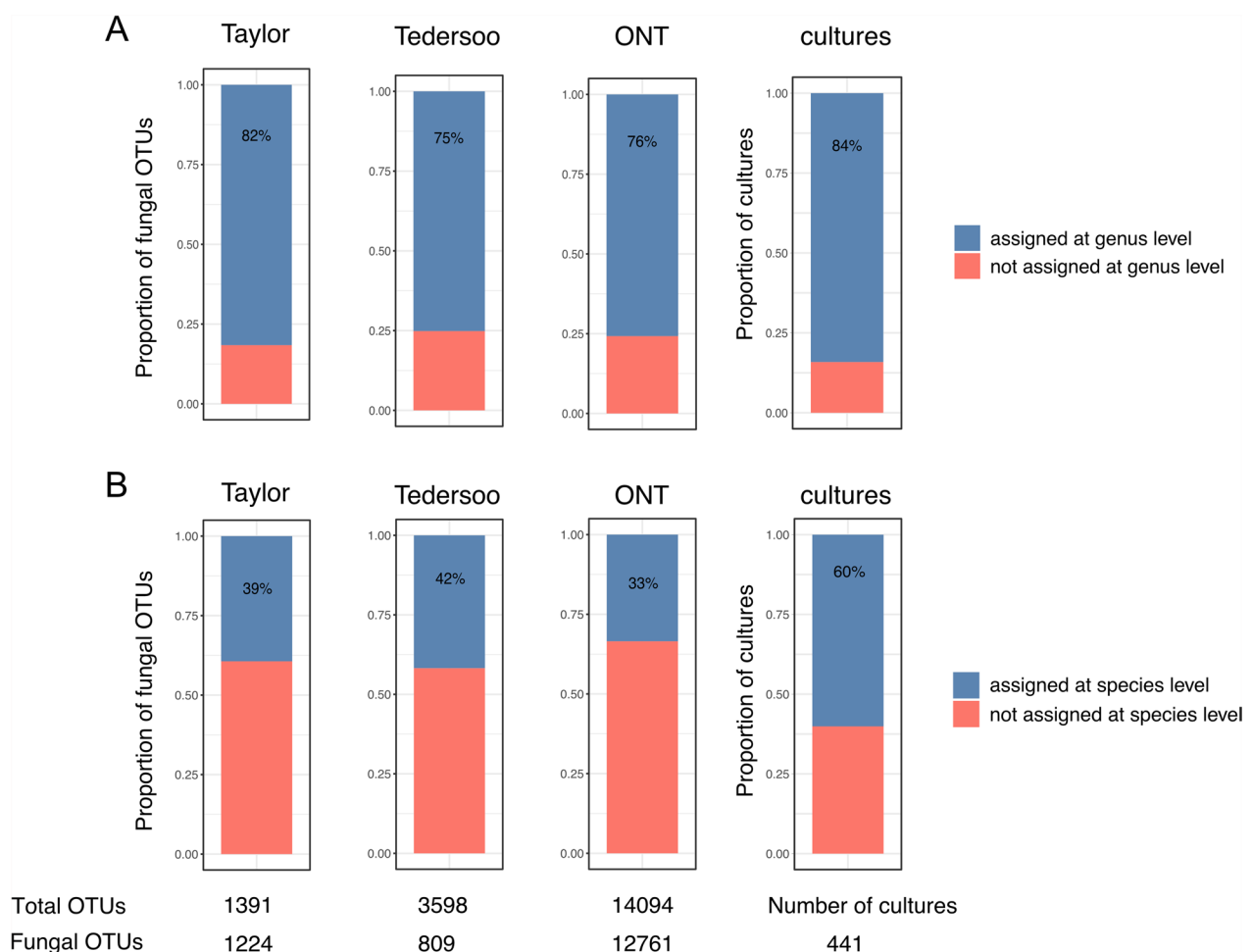
3,598 OTUs, 11% and 66% could be assigned to plants in the Taylor and Tedersoo dataset, respectively. While the proportion of reads assigned to plants in the Taylor dataset was similar in angiosperms and gymnosperms, the Tedersoo dataset contained 44% of plant reads in angiosperm samples but 88% of plant reads in gymnosperm samples (Supplementary Fig. S2). A total of 1,224 (88%) and 809 (22.5%) fungal OTUs were found in the Taylor and Tedersoo dataset, respectively. From the fungal OTUs, around 82% were assigned to genera (39% to species) in the Taylor dataset and 75% to genera (42% to species) in the Tedersoo dataset (Fig. 1). Specifically, the numbers of fungal genera identified in the Taylor dataset was 244 (253 species) while 217 genera and 234 species were identified in the Tedersoo dataset.

The bioinformatic processing of the raw long-read ONT sequences resulted in 2.42 M reads. From these

reads, 14,094 OTUs (all singletons) were obtained, of which 12,761 (90.5%) were fungal OTUs. From these fungal OTUs, 76% were assigned to 226 unique genera and 33% to 282 unique fungal species, respectively (Fig. 1).

The culturing approach identified a total of 441 fungal morphotypes based on morphological grouping of more than 4,000 isolates [6]. Based on Sanger sequencing of the ITS region, 371 morphotypes (84%) could be assigned to 62 unique fungal genera and 265 morphotypes (60%) to 72 unique fungal species (Fig. 1).

Similar results regarding the proportion of fungal OTUs assigned to genera and species were also observed when abundances (read counts) of the OTUs were considered. The relative abundances of reads belonging to fungal OTUs that were assigned to genera/species were similar across the datasets, with the Taylor dataset showing the highest proportions of fungal reads assigned



**Fig. 1** Proportions of fungal OTUs or cultures assigned to genera and species. The proportions of fungal OTUs from the Taylor, Tedersoo and ONT dataset or from the sequenced cultures that were assigned to **A** genera and **B** species. All sequences were matched to the same version of UNITE for taxonomic assignments. The total number of OTUs and number of fungal OTUs are indicated below the bars for each metabarcoding method, while the number of sequenced fungal cultures is indicated for the culturing dataset



to genera/species (84%/40%), and Tedersoo and ONT datasets following with similar values (79%/36% and 75%/37%, respectively).

In addition to summarizing fungal taxa across datasets, this study also investigated the distribution of fungi into different primary lifestyles (i.e., potential plant pathogens, plant saprotrophs or plant endophytes) based on their genus-level assignments. In the Taylor dataset, 39% of the fungal OTUs (480 out of 1,224) were assigned to one of the above-mentioned lifestyles, while around 46% of the fungal OTUs (373 out of 809) were assigned to a lifestyle in the Tedersoo dataset (Supplementary Table S4). Around 44% of fungal OTUs were assigned to a lifestyle in the ONT dataset (5,621 out of 12,761 OTUs), as well as in the culturing dataset (179 out of 441 morphotypes).

#### Fungal community comparison between metabarcoding and culturing datasets

A comparison of fungal genera identified from each of the four methods showed the largest overlaps across the three sequencing-based methods (Fig. 2A). Almost 40% of all identified genera (122 out of 311 overall identified genera, Fig. 2A) were found exclusively in the three sequencing-based datasets but not in the culturing dataset. The Taylor and Tedersoo datasets shared over 80% of their total genera, and each of them shared around 60% of the genera with the ONT dataset. Only 39 out of 311 genera (12.5% of all identified genera) were shared across all four datasets. Portions of dataset-specific genera in the metabarcoding datasets ranged from approximately 3% (8 out of 217 genera, Tedersoo dataset) to 12% (36 out of 226 genera, ONT dataset), and they belonged to all three investigated fungal lifestyles (Additional file 1). Noteworthy, ten genera (out of 62 cultured genera) were exclusively found by culturing (Additional file 1).

A direct comparison of sequences obtained from the three metabarcoding datasets against the sequences obtained by culturing revealed that more than 80% of the reads in the Taylor (88%) and Tedersoo (83%) dataset matched against the cultured fungi, while more than half of the reads of the ONT dataset could not be matched to sequences of the cultures but were assigned to species within the UNITE database (65%). The metabarcoding communities differed in their overlap with cultured sequences not only between the datasets, but also in dependency of the host tree group or even the sample (Supplementary Fig. S3).

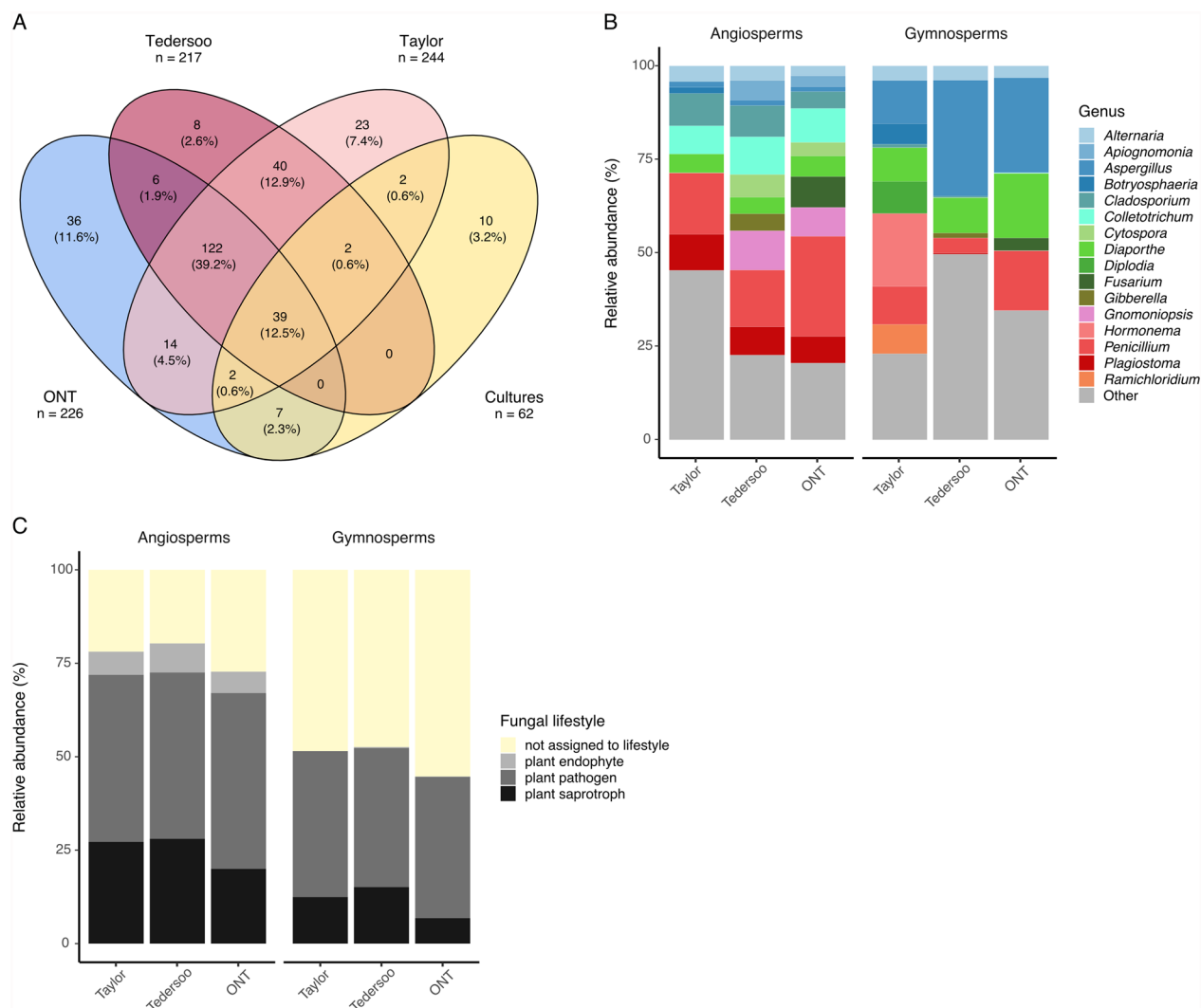
Comparing the fungal communities between metabarcoding methods revealed that *Alternaria*, *Aspergillus*, *Colletotrichum*, *Diaporthe*, *Penicillium* and *Plagiostoma* were recorded among dominant genera in all three datasets (Fig. 2B). Some genera appeared in all datasets and

both host tree groups in relatively high abundances, for example *Diaporthe*, which on average amounted to more than 4% in angiosperm seeds and more than 6% in gymnosperm seeds. At the same time, some of the genera dominant across datasets were differently abundant in angiosperms and gymnosperms, as indicated by the enrichment analysis (Supplementary Table S3). For example, the average relative abundance of *Colletotrichum* was higher in angiosperms than gymnosperms in all three datasets. *Plagiostoma*, *Cladosporium* and *Gnomoniopsis* were also more relatively abundant in angiosperms than gymnosperms (Fig. 2B), but this trend was significant only in the Taylor and Tedersoo dataset. *Penicillium* showed a similar trend in all three datasets, but differential abundances between host tree groups were significantly different only in the Tedersoo dataset. The opposite trend, i.e., higher relative abundances in the gymnosperm over angiosperm seeds, was observed for *Diplodia*, *Hormonema* and *Aspergillus* in all three datasets, but differences in *Aspergillus* were only significant in the Taylor and ONT, and in *Hormonema* only in the Taylor dataset (Supplementary Table S3). In dependency of host tree group and dataset, *Fusarium* was among the dominant genera only in the ONT dataset and enriched, although not significantly, in angio- vs. gymnosperms, while *Ramichloridium* was among the twelve dominant genera only in the Taylor dataset (Fig. 2B) and more frequent in gymnosperms than in angiosperms (Supplementary Table S3).

Of the genus-level assigned OTUs that were also assigned to a lifestyle, plant pathogens had the highest number of OTUs and OTU read abundances, followed by plant saprotrophs and endophytes (Supplementary Table S4). Relative abundances of the lifestyles were similar across datasets but varied to a certain extent between host tree groups (Fig. 2C). Endophytes overall were more prevalent in angiosperm seeds than in gymnosperm seeds (enrichment analysis:  $P < 0.05$ ) in all datasets, although endophyte abundances varied among angiosperm samples. For example, unlike other angiosperm seeds, *Quercus garryana* seeds harbored endophytes in high relative abundances (>30%) that primarily belonged to *Gnomoniopsis* sp. or *G. paraclavulata* in all datasets.

#### Fungal alpha-diversity comparison between host tree groups and metabarcoding datasets

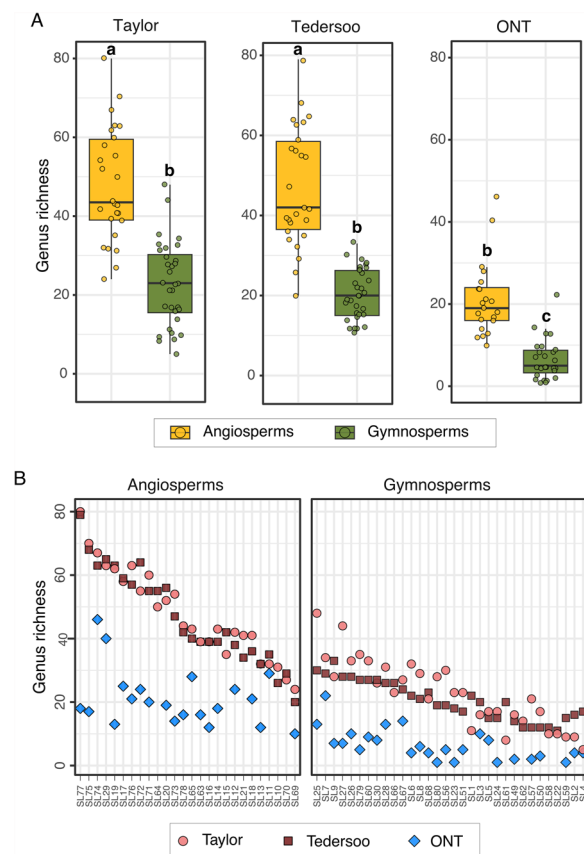
A higher richness of fungal genera in angiosperm in comparison with gymnosperm seed samples ( $\chi^2 = 39.01$ ,  $df = 1$ ,  $P < 0.001$ ) was revealed (Fig. 3). However, genus richness was similar in the Taylor and Tedersoo datasets and higher than in ONT dataset ( $\chi^2 = 97.00$ ,  $df = 2$ ,  $P < 0.001$ ; Fig. 3A). Although the interaction term between host tree group and method



**Fig. 2** Differences in genus composition and lifestyles of the seed mycobiota across methods. **A** The overlap of unique genera identified across the Taylor, Tedersoo, ONT metabarcoding and culturing datasets. The numbers inside the Venn Diagram are genera counts and proportions related to the overall (total) number of genera identified across all methods ( $n=311$ ), while the total numbers of genera identified with each method are indicated outside of the diagram. **B** Community profiles of fungi identified in the seeds of the two host tree groups (angio- and gymnosperms), in the three metabarcoding datasets. A total of 26 angiosperm and 32 gymnosperm seed lots were analyzed for the Taylor and Tedersoo datasets, respectively, while 21 angiosperm and 27 gymnosperm seed lots were analyzed in the ONT dataset. The bar plot shows average relative abundances of taxa assigned at genus level, across samples within each host tree group. The twelve most abundant genera identified for each of the datasets were used to generate the plot. Thus, some genera with small average values in either host tree group may not be easily visible but are still present among the overall dominating genera within each dataset. Genera not among the top twelve taxa were summed up and labeled as 'Other'. **C** The distribution of fungi identified at genus level into lifestyles and their average relative abundances in the seeds of the two host tree groups (angio- and gymnosperms) for the three metabarcoding datasets

was significant ( $\chi^2 = 10.36$ ,  $df=2$ ,  $P < 0.05$ ), indicating a method-dependent trend of differences in genus richness between angiosperms and gymnosperms, this was not observed with post hoc testing (Fig. 3). Per-sample richness of genera was on average half as low in the ONT dataset compared to both Illumina metabarcoding datasets, that is, on average 47.5 (angiosperms) and 21.6 (gymnosperms) genera were identified per sample

in the Taylor and Tedersoo datasets, while 21.1 (angiosperms) and 6.5 genera (gymnosperms) were identified per sample by ONT sequencing. Fungal OTU richness was higher in the ONT than in the Taylor and Tedersoo datasets ( $\chi^2 = 53.80$ ,  $df=2$ ,  $P < 0.001$ ), and in angiosperm than in gymnosperm samples ( $\chi^2 = 24.68$ ,  $df=1$ ,  $P < 0.001$ ) in all three dataset (as indicated by a non-significant interaction between host tree group

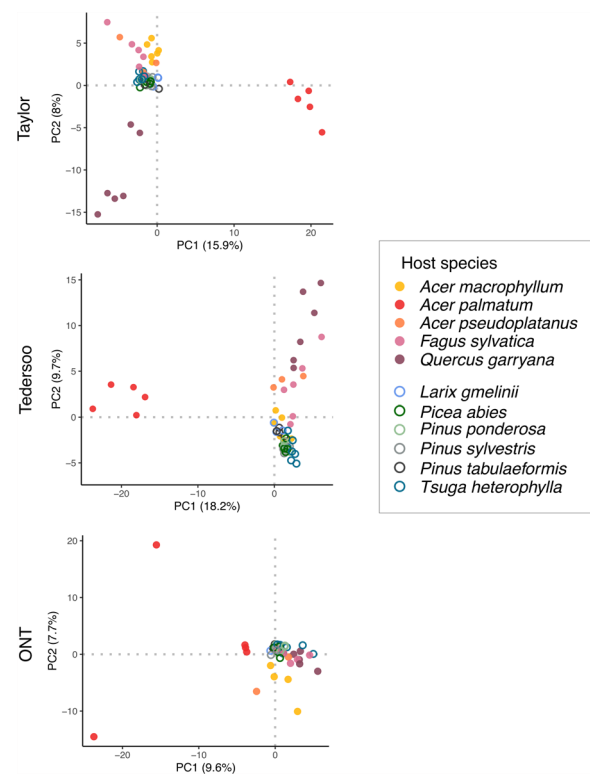


**Fig. 3** Alpha-diversity of seed mycobiota across host tree groups and methods. **A** Genus richness in the Taylor, Tedersoo and ONT datasets, in the angio- and gymnosperm seed fungal communities. Boxplots show the median and interquartile ranges for the plotted measures, and each point represents one sequenced sample. Different letters above the bars indicate significantly different values ( $P < 0.05$ ). **B** Per-sample genus richness across samples belonging to different host tree groups and methods ordered based on the number of fungal genera per sample

and method;  $\chi^2 = 0.96$ ,  $df = 2$ ,  $P = 0.62$ ; Supplementary Fig. S4A).

Based on abundance-weighted alpha-diversity measures (i.e., Shannon's entropy and the Inverse Simpson Diversity) and the relative abundance of the most dominant OTU for the Taylor and Tedersoo datasets (Supplementary Fig. S4B-C), lower evenness was found in the gymnosperm than angiosperm communities in both datasets, indicating that the angiosperm seed communities consisted of a higher number of more evenly abundant OTUs than gymnosperms, which were dominated by a lower number of OTUs.

Considering fungal lifestyles, within each dataset, angiosperms overall hosted a slightly higher diversity of plant pathogenic genera and saprotrophic genera than gymnosperms (Supplementary Fig. S5). Plant endophytes were the least genus-diverse group in all three datasets



**Fig. 4** The structure of the mycobiota in seed samples from different host tree species. PCA plots are based on fungal genera revealed by the Taylor, Tedersoo and ONT dataset and their CLR-normalized abundances. The samples are colored according to host tree species, with filled points representing angiosperm and hollow points representing gymnosperm samples. The full set of genera identified within each dataset were used for the genus-level PCA analyses (Taylor:  $n = 244$ , Tedersoo:  $n = 217$ , ONT:  $n = 226$ )

and in both host groups. In addition, not all seed samples harbored genera with an endophytic lifestyle, and seed lots without plant endophytic genera were more often found in gymnosperms (Supplementary Fig. S5). A joint analysis across methods of the per-sample proportions of genera assigned to different lifestyles showed that the combination of method and host group significantly influenced the proportions of genera in the different lifestyles ( $P < 0.01$ ,  $df = 8$ ).

#### Fungal beta-diversity comparison between host tree groups and metabarcoding datasets

PCA analyses were conducted to visually determine whether the fungal communities were structurally clustered based on host tree group and host tree species (indirectly related to their continent of origin). Fungal community structure at genus level in all three metabarcoding datasets revealed similar patterns related to host tree group and host tree species in Taylor and Tedersoo dataset (Fig. 4). For example, in addition to communities

from the North American oak *Q. garryana*, communities from *Acer palmatum* repeatedly clustered apart from the main bulk of communities in the Taylor and Tedersoo dataset. In the ONT dataset, samples belonging to *A. palmatum* were also separated from the rest of the communities along PC1, but to a less striking extent. Overall, the proportion of variance explained by the first two PCs was higher in both Taylor (23.9%) and Tedersoo (27.9%) datasets in comparison with the less differentiated ONT data (17.3%). In all datasets, the gymnosperm seed samples did not spread along the axes as much as the angiosperm samples. Similar to fungal communities at genus level, the communities at OTU level in the Taylor and Tedersoo dataset showed separation based on a combination of host tree group and species (Supplementary Fig. S6). The communities of the Asian *A. palmatum* clearly differed from most of the other communities along PC2 in both datasets. The Taylor dataset further emphasized structural differences in communities of the North American *Acer macrophyllum*, while communities of the North American *Pinus ponderosa* separated from the bulk of communities in the Tedersoo dataset on PC1.

Similar results were obtained with PERMANOVA, which revealed that host tree group and method, as well as their interaction, had significant effects on the community composition of fungal genera (Table 2). However, only about 10% of variation in fungal communities across samples was explained by those factors, with host tree group alone accounting for 6% of the variation (Table 2). Structuring of the fungal communities was also observed at a host tree species level (Fig. 4)—the PERMANOVA model containing host tree species, method, and their interaction, in which all explanatory variables were significant (Table 2), explained almost 50% of variance with host tree species accounting for 33% alone. Furthermore, pairwise comparisons based on host tree group and method showed that the Taylor and Tedersoo datasets revealed similar fungal communities, which showed

separation based on host tree group (angiosperms vs. gymnosperms; Supplementary Table S5). Differences between angio- and gymnosperm communities were also noticed in the ONT dataset, but those were different from the communities of angiosperm and gymnosperm trees in the Taylor and Tedersoo dataset (Supplementary Table S5).

These structural patterns were corroborated by community-level analyses of abundances in heatmaps, which highlighted clusters of genera that were consistently, that is, across all three metabarcoding datasets, differentially abundant in specific subsets of samples (Supplementary Fig. S7). For example, seven fungal genera formed a separate cluster of highly abundant genera in *A. palmatum* in both the Taylor and Tedersoo dataset (and *Pseudopithomyces* as additional genus in the Tedersoo dataset). Although five of those genera were also abundant in the ONT samples of *A. palmatum*, they did not form a separate cluster, and distinct genera were additionally identified as abundant. Both *Gnomoniopsis* and *Cytospora* were abundantly found in samples of North American *Q. garryana* tree seeds in all three datasets, as well as *Neocucurbitaria*, *Apiognomonina* and *Ramularia* in *Fagus sylvatica*. In certain host tree species, each dataset revealed further abundant genera: for *F. sylvatica*, *Phoma* was abundantly found in the Taylor and Tedersoo dataset, *Seimatosporium* in the Tedersoo and ONT dataset, and *Didymocyrtys* in the ONT dataset exclusively. Interestingly, community analyses revealed a higher level of homogeneity within the gymnosperm seeds in all three methods: samples from gymnosperm seeds generally clustered more closely together irrespective of host tree species or origin and shared more fungal genera, while angiosperm samples belonging to different host tree species or origins were distinct from each other (Fig. 3, Supplementary Figs. S6 and S7).

## Discussion

Despite the high diversity of fungal OTUs and genera revealed by metabarcoding approaches in our study, a substantial number of reads and OTUs could not be assigned to genera or species, and, contrary to our expectations, this applied to both short- and long-read metabarcoding. Third-generation sequencing technologies, to which ONT belongs, produce reads of up to hundreds of kilobases vastly exceeding the lengths that can be obtained by Illumina sequencing. Unlike other studies which found higher fractions of reads assigned to fungal species in ONT vs. Illumina metabarcoding [50], we could not confirm that more reads or OTUs were identified to species in our full-length ITS ONT vs. ITS2 Illumina datasets. One partial explanation for this result may be the general lack of fungal type strain reference

**Table 2** Results of PERMANOVAs testing for differences in community structure of the mycobiota of seeds

| Factors               | DF | F      | R <sup>2</sup> | P-value      |
|-----------------------|----|--------|----------------|--------------|
| Method                | 2  | 1.610  | <b>0.018</b>   | <b>0.007</b> |
| Host group            | 1  | 10.925 | <b>0.063</b>   | <b>0.001</b> |
| Method * host group   | 2  | 1.742  | <b>0.020</b>   | <b>0.004</b> |
| Method                | 2  | 2.205  | <b>0.018</b>   | <b>0.001</b> |
| Host species          | 10 | 7.790  | <b>0.326</b>   | <b>0.001</b> |
| Method * host species | 20 | 1.343  | <b>0.112</b>   | <b>0.001</b> |

The results are based on genus-level community matrices for which the robust Aitchison distances were calculated. Degrees of freedom (DF), R<sup>2</sup> and F values are shown. Significant P-values ( $P < 0.05$ ) and corresponding R<sup>2</sup> values are given in bold



sequences and a high estimated fraction of undescribed fungi [96]. Although the number of entries of full-length ITS sequences in reference databases is increasing [97], there is a need to continue generating sequences of previously and newly described species. On one hand, this can be accomplished by intensifying and optimizing culturing and taxonomic classification to improve taxonomy-based assignments. On the other hand, sequence-based phylogenetic placement of fungi may become an important alternative for overall fungal identification [51, 54]. These two approaches should not be mutually exclusive. Another partial cause for the limited success in OTU assignments to genera and species in our ONT dataset may be related to the high total error rates (approx. 7.5%) characteristic of the ONT version (R9.4.1) [66, 67] that was used in our study. Recently, ONT error rates have been significantly reduced [47, 98] which, together with the continuous expansion of reference databases [26] and a better understanding of ONT sequence quality [99] to develop streamlined processing options, will likely result in higher taxonomic resolution of long ONT reads.

#### Primer selection

Although there was a substantial overlap of the fungal genera recovered by short- and long-read metabarcoding, shifts in the ranking of genera according to their relative abundance were observed between methods and hosts. For example, the genus *Fusarium* was identified among the top twelve genera in the ONT dataset, while it was not among the top genera in the two Illumina metabarcoding datasets. Such discrepancies may be due to the preferential amplification of certain taxa by different primers depending on the part of the ITS region sequenced [39, 42, 100]. Therefore, metabarcoding primers should be optimized to ensure the detection of the target taxa [36, 53, 57, 101, 102], and ideally, multiple primer pairs should be used to ensure the amplification of the broader fungal community.

Primer selection is critical not only because different primer pairs may target different marker genes, resulting in amplification of distinct fungal community profiles [103], but also because primers may affect the depth and thus resolution of a metabarcoding assay due to co-amplification of host (plant) DNA. In this study, the *ITS3ngs-mix/ITS4ngsUni* primer pair used in the Tedersoo dataset generated 2.49 million sequences, of which 1.63 million (65%) were plant reads. The use of non-specific primers and co-amplification of plant reads in metabarcoding studies may reduce the probability of capturing all fungi in sequenced samples, especially those belonging to rare fungal taxa. However, our results show that the Tedersoo data consisted of similar fungal genera as the Taylor dataset. This implies that the sequencing

depth was sufficient to capture most of the fungal diversity despite a large fraction of reads being 'wasted' on plant sequences, which was especially the case for certain gymnosperm samples. Thus, reducing plant co-amplification by selecting primers specific for fungi or using peptide nucleotide blockers to block host contamination when using non-specific primers [104, 105], can lead to a more complete assessment of fungal diversity.

#### Fungal culturing

Culture-free (i.e., metabarcoding) approaches revealed more than three times the number of fungal genera compared to the culture-based approach. Greater richness in metabarcoding than culturing datasets has been repeatedly observed in previous studies of plant mycobiomes [6, 106, 107] and likewise in oomycete microbiomes [108]. However, in our study most of the dominant taxa from the Illumina datasets were also cultured, as it was previously observed for foliar fungi of *F. sylvatica* [109], suggesting that higher observed diversities in metabarcoding datasets partially stem from rare taxa. A unique feature of the culture-based approach is that only living and thus metabolically active microbes, except obligate biotrophs, can be cultured. Hence, a smaller number of taxa can be expected to be identified with this method compared to metabarcoding, which also detects relic DNA [110], DNA belonging to metabolically inactive organisms, and to obligate biotrophs. Furthermore, in our study, several fungal taxa were observed uniquely in the culturing dataset, similar as previously described by Oita et al. [106], which suggests that culturing and metabarcoding should be used simultaneously if possible. Since pure cultures are indispensable for follow-up experiments related to fungal trait assessment or interactions with host plants or other organisms, additional effort in this area could improve culture-dependent diversity assessments. Some of the possibilities include cutting plant tissues into small pieces or even grinding them before placing them on agar media to reduce competition among fungi and to allow rare and slow-growing fungi to grow [23, 111]. In addition, different growing conditions and multiple nutrient media, including those supplemented with plant material, can be used to ensure that a wide range of endophytes can be captured [21].

#### Comparison of metabarcoding datasets

The ONT dataset revealed lower genus richness and a different fungal community structure than the Taylor and Tedersoo datasets. As previously described by Furneaux et al. [51], the higher frequency of rare taxa due to greater per-sample sequencing depth in short-read sequencing compared to long-read sequencing can contribute to lower per-sample genus diversity in long-read datasets,

and this was likely the case in our study. A large amount of reads originally generated by the ONT sequencing was discarded with the bioinformatic pipeline, in addition to erroneous sequences and disregarding other initially sequenced markers (SSU and LSU). These processing steps, together with the high error rates which were discussed before, likely contributed to the reduced number of genera identified per sample. These limitations should be considered when samples with high expected diversity (including rare taxa) are studied, which may have applied to the angiosperm samples of our study. In our case, the discrepancy of observed genus richness between the Illumina and ONT datasets was more prominent for the angiosperm samples than for the gymnosperm samples, which was likely due to differences in observed proportions of rare OTUs in these two groups of samples and their relative contribution to fungal diversity. Interestingly, a study using ONT and Illumina to characterize fungi associated with tall fescue [50] found a more diverse community and more rare taxa in the ONT dataset, stressing that no absolute agreements exist regarding the outcome of using ONT across studies. Loss of rare taxa may also affect beta-diversity patterns – in our study, short-read metabarcoding revealed more structured communities along host tree species than long-read metabarcoding. This result aligns with our expectation that ONT, given its lower sequencing depth and higher error rates, will have reduced diversity coverage power compared to Illumina sequencing. ONT sequencing with newer MinION flow cells can be used to detect specific pathogenic fungi [47], which has proven useful in the field of diagnostics, where rapid approaches and long reads are needed [112, 113]. For diagnostic purposes, it should however be clear that important considerations about sensitivity (detection limits) and quantitative reliability would have to be made [55].

#### Ecological patterns across metabarcoding datasets

Despite the differences in genus richness and fungal community composition between short- and long-read metabarcoding, higher richness of fungal genera in angiosperm seeds compared to gymnosperm seeds was observed across datasets. The trend of higher alpha-diversity in angiosperms than gymnosperms was previously observed in a global study of tree seeds from botanical gardens [4]. Given that sexual reproduction, seed development and seed structures show remarkable discrepancies between angiosperms and gymnosperms [114], the variation in fungal diversity as a function of host plant group is not surprising. This study also revealed differences in fungal community composition between host tree groups and host tree species, and these

host-dependent patterns were visible in all three datasets. While the gymnosperm samples clustered more closely together in the PCAs and did not create distinct clusters in the heatmaps, the fungal communities among the angiosperm samples showed the opposite tendencies, indicating that tree species-dependent community structure was stronger in angiosperms than in gymnosperms. Phylogenetic proximity of the hosts and host functional traits were previously shown to influence microbial composition of fungal endophytes in tropical seeds [115], tree twigs [116], leaves [107, 117], roots [118] and bark [119], suggesting co-evolution of plants and their associated mycobiomes. The breadth of angiosperm clades and the evolutionary success of flowering plants [120] may thus have contributed to the concurrent explosion of fungi associated with them, while the phylogenetically older and less divergent gymnosperm tree species may instead host a more similar mycobiota.

Besides comparing the fungal taxonomic diversity revealed by different methods, our study also examined potential fungal lifestyles (i.e., plant pathogens, plant saprotrophs and plant endophytes) captured by different methods. Currently, little is known about the function of most seed-borne fungi, but primary lifestyle assignments can be made based on genus-level taxonomy using tools such as FUNGuild [71] or FungalTraits [70], as functional traits are often conserved at the genus or sometimes even higher taxonomic level. However, a large proportion of OTUs in metabarcoding studies cannot be assigned to a genus and therefore to a lifestyle, or lifestyle information is simply missing. This problem is amplified by the previously highlighted structure of reference databases commonly used for taxonomic assignment, with entries traditionally stemming from cultured type strains. Unculturable taxa or taxa without reference sequences, at the moment, cannot be taxonomically placed (i.e., so-called ‘dark’ taxa [121]) and thus cannot be used to further decrypt potential function. For example, in our study, approximately 50% of fungal OTUs were assigned to one of the three targeted lifestyles in all metabarcoding datasets, similar to a previous study in which half of the fungal OTUs obtained from leaves and needles of several tree species were assigned to a lifestyle [122].

Of the fungal OTUs assigned to a lifestyle, plant endophytes were less abundant in the seed samples than plant saprotrophs and plant pathogens, respectively, and this was observed across datasets. In our study, around 3% of the genera identified in each dataset were assigned to plant endophytes. Partly, this pattern may be explained by a low number of endophyte records in the FungalTraits database [70], i.e., only around 1% of more than 10,000 entries are assigned to plant endophytes. One reason for

the lack of fungal endophytes in the database could also be due to their biotrophic nature, making them difficult to isolate, describe and name [123], and to a historical research bias towards plant pathogens. Finally, fungal lifestyles vary from pathogenic to mutualistic depending on numerous abiotic and biotic factors that may change over time [124]. Thus, the taxonomy-based lifestyle assignments of our samples represent only a snapshot of all potential lifestyles the mycobiota can have. To better understand fungi-host interactions, more research is needed. This could include metagenomic or metatranscriptomic experimental studies to assess the functional potential or active genes of entire mycobiome communities [112], for which new long-read sequencing technologies can be a useful approach.

## Conclusions

Here, we used metabarcoding (short-read and long-read sequencing) and culturing to assess fungi from the same seed samples belonging to several tree species. The datasets were compared to understand if similar fungal communities were recovered by all employed methods. Across the three metabarcoding datasets, similar absolute numbers of fungal genera and species were revealed. About half of the characterized fungi overlapped, but each dataset also identified unique taxa, depending on the host tree group and species. The discrepancies in taxonomic composition between metabarcoding methods may be indirectly related to processing steps (e.g., primer selection and bioinformatic processing) and sequence length, but also the relatively higher error rate in the ONT dataset due to old chemistry. This may have also influenced the higher relative fraction of reads in the Illumina datasets that were assigned to cultures compared with the ONT dataset. Despite these differences, similar host-dependent alpha- and beta-diversity patterns and functional lifestyle compositions were revealed. For biodiversity studies, Illumina metabarcoding might be a suitable method to use thanks to its high diversity coverage, affordable price and availability of established processing pipelines. Along with ongoing reductions of error rates and the population of reference sequence databases, we see opportunities of long-read ONT sequencing as a tool for diagnostic purposes, where the user is interested in confirming known species. Higher accuracies and sufficient sequencing depths provided, ONT is also advancing in community analysis. Finally, for research aiming at assessing fungal traits, the sequencing and archiving of pure cultures are still indispensable.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-025-00712-7>.

Additional file 1.

Additional file 2.

Additional file 3.

## Acknowledgements

We acknowledge the Genetic Diversity Centre (ETH Zurich) for the bioinformatic support and providing computational infrastructure. We thank the four anonymous reviewers for helpful comments on the manuscript.

## Author contributions

IF, SS, SP and RE designed the study. IF coordinated the sampling. Lab protocols were prepared by QK. Lab work was done by IF (culturing and Illumina metabarcoding) and QK (ONT metabarcoding) with support from SS. MH bioinformatically processed the Illumina raw data and RH and QK bioinformatically processed the ONT raw data. JM and IF analyzed the data and wrote the manuscript. JM and IF finalized the manuscript with comments from the other co-authors.

## Funding

This work was partially funded by the Swiss National Science Foundation SNSF (Grant: 174644) and by the Swiss Federal Office for the Environment (FOEN) (Finanzhilfvertrag betreffend wissenschaftlich-technischen Tätigkeiten im Bereich Waldschutz). René Eschen was supported by CABI. CABI is an international intergovernmental organisation, and we gratefully acknowledge the core financial support from our member countries (and lead agencies) including the UK (Foreign, Commonwealth & Development Office), China (Chinese Ministry of Agriculture and Rural Affairs), Australia (Australian Centre for International Agricultural Research), Canada (Agriculture and Agri-Food Canada), Netherlands (Directorate-General for International Cooperation) and Switzerland (Swiss Agency for Development and Cooperation). See <https://www.cabi.org/about-cabi/who-we-work-with/key-donors/> for full details. The Nanopore MinION and ONT consumables were purchased with WSL internal grant PID-1718(1).

## Data availability

Sequencing data from this study is deposited at ENA under accession numbers PRJEB81281 (Tederloo) and PRJEB81455 (ONT), and data previously generated by Franić et al. (2019) under PRJNA550270 (Taylor) and MN105153–MN105593 (cultured sequences). The scripts required for data analysis and plotting, and the metadata are deposited at bitbucket ([https://bitbucket.org/janami/tree\\_seed\\_metabarcoding\\_methods](https://bitbucket.org/janami/tree_seed_metabarcoding_methods)).

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft WSL, Birmensdorf, Switzerland. <sup>2</sup>CABI, Delémont, Switzerland. <sup>3</sup>Institute of Agricultural Sciences, ETH Zürich, Zurich, Switzerland.

Received: 31 October 2024 Accepted: 20 April 2025

Published online: 19 May 2025

## References

- Simonin M, Briand M, Chesneau G, Rochefort A, Marais C, Sarniguet A, et al. Seed microbiota revealed by a large-scale meta-analysis including 50 plant species. *New Phytol.* 2022;234:1448–63.
- Latz MAC, Kernn MH, Sørensen H, Collinge DB, Jensen B, Brown JKM, et al. Succession of the fungal endophytic microbiome of wheat is dependent on tissue-specific interactions between host genotype and environment. *Sci Total Environ.* 2021;759:143804.
- Fort T, Pauvert C, Zanne AE, Ovaskainen O, Caignard T, Barret M, et al. Maternal effects shape the seed mycobiome in *Quercus petraea*. *New Phytol.* 2021;230:1594–608.
- Franić I, Eschen R, Allan E, Hartmann M, Schneider S, Prospero S. Drivers of richness and community composition of fungal endophytes of tree seeds. *FEMS Microbiol Ecol.* 2020;96:fiaa166.
- Cleary M, Oskay F, Doğmuş HT, Lehtijärvi A, Woodward S, Vettriano AM. Cryptic risks to forest biosecurity associated with the global movement of commercial seed. *Forests.* 2019;10:459.
- Franić I, Prospero S, Hartmann M, Allan E, Auger-Rozenberg M-A, Grünwald NJ, et al. Are traded forest tree seeds a potential source of nonnative pests? *Ecol Appl.* 2019;29:e01971.
- Franić I, Eschen R, Allan E, Hartmann M, Schneider S, Prospero S. Drivers of richness and community composition of fungal endophytes of tree seeds. *FEMS Microbiol Ecol.* 2020;96:1–10.
- Wentzien NM, Fernández-González AJ, Valverde-Corredor A, Lasa AV, Villadas PJ, Wicaksono WA, et al. Pitting the olive seed microbiome. *Environ Microbiome.* 2024;19:17.
- Abdelfattah A, Wisniewski M, Schena L, Tack AJM. Experimental evidence of microbial inheritance in plants and transmission routes from seed to phyllosphere and root. *Environ Microbiol.* 2021;23:2199–214.
- Franić I, Cleary M, Aday Kaya AG, Bragança H, Brodal G, Cech TL, et al. The biosecurity risks of international forest tree seed movements. *Curr For Rep.* 2024;10:89–102.
- Kumar R, Gupta A. Seed-borne diseases of agricultural crops: Detection, diagnosis & management. Singapore: Springer; 2020.
- Storer AJ, Gordon TR, Clark SL. Association of the pitch canker fungus, *Fusarium subglutinans f.sp. pini*, with Monterey pine seeds and seedlings in California. *Plant Pathol.* 1998;47:649–56.
- Decourcelle T, Piau D, Desprez-Loustau ML. Detection of *Diplodia sapinea* in Corsican pine seeds. *Plant Pathol.* 2015;64:442–9.
- Burgess T, Wingfield MJ. Quarantine is important in restricting the spread of exotic seed-borne tree pathogens in the southern hemisphere. *Int For Rev.* 2002;4:56–65.
- Delgado-Sánchez P, Jiménez-Bremont JF, Guerrero-González ML, Flores J. Effect of fungi and light on seed germination of three *Opuntia* species from semiarid lands of central Mexico. *J Plant Res.* 2013;126:643–9.
- Rétif F, Kunz C, Calabro K, Duval C, Prado S, Bailly C, et al. Seed fungal endophytes as biostimulants and biocontrol agents to improve seed performance. *Front Plant Sci.* 2023;14:1260292.
- Baldwin TT, Zitomer NC, Mitchell TR, Zimeri AM, Bacon CW, Riley RT, et al. Maize seedling blight induced by *Fusarium verticillioides*: accumulation of fumonisin B1 in leaves without colonization of the leaves. *J Agric Food Chem.* 2014;62:2118–25.
- Li YM, Shaffer JP, Hall B, Ko H. Soil-borne fungi influence seed germination and mortality, with implications for coexistence of desert winter annual plants. *PLoS ONE.* 2019;14:e0224417.
- Bergmann GE, Heitmann SJ, Busby PE, Leveau JHJ. Characterization of seed mycobiota using culture-dependent and culture-independent approaches. In: *Methods in molecular biology*. Humana Press Inc.; 2023. p. 65–78.
- Luchi N, editor. *Plant Pathology—Methods and Protocols*. New York: Springer, US; 2022.
- dos Reis JBA, Lorenzi AS, do Vale HMM. Methods used for the study of endophytic fungi: a review on methodologies and challenges, and associated tips. *Arch Microbiol.* 2022;204:675.
- Spanu PD. The genomics of obligate (and nonobligate) biotroph. *Annu Rev Phytopathol.* 2012;50:91–109.
- Unterseher M, Schnittler M. Dilution-to-extinction cultivation of leaf-inhabiting endophytic fungi in beech (*Fagus sylvatica* L.)—different cultivation techniques influence fungal biodiversity assessment. *Mycol Res.* 2009;113:645–54.
- Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol.* 2016;34:942–9.
- Knight R, Vrbanac A, Taylor BC, Aksenov A, Callewaert C, Debelius J, et al. Best practices for analysing microbiomes. *Nat Rev Microbiol.* 2018;16:1–13.
- Abarenkov K, Nilsson RH, Larsson KH, Taylor AFS, May TW, Frøslev TG, et al. The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. *Nucleic Acids Res.* 2024;52:D791–7.
- Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). National Human Genome Research Institute NIH. 2020. <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>. Accessed 22 May 2021.
- Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature.* 2012;486:222–7.
- Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, et al. Microbial interkingdom interactions in roots promote arabidopsis survival. *Cell.* 2018;175:973–983.e14.
- Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ. Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc Natl Acad Sci.* 2018. <https://doi.org/10.1073/pnas.1717617115>.
- Coleman-Derr D, Desgarnies D, Fonseca-Garcia C, Gross S, Clingenpeel S, Woyke T, et al. Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytol.* 2016;209:798–811.
- Thiergart T, Duran P, Ellis T, Garrido-Oter R, Kemen E, Roux F, et al. Root microbiota assembly and adaptive differentiation among European *Arabidopsis* populations. *Nat Ecol Evol.* 2019;12:640623.
- Tedersoo L, Bahram M, Pölme S, Kõljalg U, Yorou NS, Wijesundera R, et al. Global diversity and geography of soil fungi. *Science.* 1979;2014(346):1256688.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. *Science.* 1979;2015(348):1261359–1261359.
- Fricker AM, Podlesny D, Fricke WF. What is new and relevant for sequencing-based microbiome research? A mini-review. *J Adv Res.* 2019;19:105–12.
- Wurzbacher C, Larsson E, Bengtsson-Palme J, Van den Wyngaert S, Svantesson S, Kristiansson E, et al. Introducing ribosomal tandem repeat barcoding for fungi. *Mol Ecol Resour.* 2018;00:1–10.
- Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, et al. Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol.* 2013;22:5271–7.
- Abarenkov K, Zirk A, Piirmann T, Põhönen R, Ivanov F, Nilsson RH, et al. UNITE QIIME release for Fungi 2. UNITE Community. 2021.
- Tedersoo L, Anslan S, Bahram M, Pölme S, Riit T, Liiv I, et al. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *Mycosyst Res.* 2015;10:1–43.
- Tedersoo L, Lindahl B. Fungal identification biases in microbiome projects. *Environ Microbiol Rep.* 2016;8:774–9.
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A.* 2012;109:6241–6.
- Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kausrud H. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiol.* 2010;10:1–9.
- Heeger F, Wurzbacher C, Bourne EC, Mazzoni CJ, Monaghan MT. Combining the 5.8S and ITS2 to improve classification of fungi. *Methods Ecol Evol.* 2019;10:1702–11.
- Yang RH, Su JH, Shang JJ, Wu YY, Li Y, Bao DP, et al. Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. *PLoS ONE.* 2018;13:e0206428.
- Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nat Rev Microbiol.* 2019;17:95–109.



46. Runnel K, Abarenkov K, Copot O, Mikryukov V, Kõljalg U, Saar I, et al. DNA barcoding of fungal specimens using PacBio long-read high-throughput sequencing. *Mol Ecol Resour.* 2022. <https://doi.org/10.1111/1755-0998.13663>.
47. Loit K, Adamson K, Bahram M, Puusepp R, Anslan S, Kiiker R, et al. Relative performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) third generation sequencing instruments in identification of agricultural and forest fungal pathogens. *Appl Environ Microbiol.* 2019;85:e01368.
48. Kirchgeorg S, Chang JJM, Ip YCA, Jucker M, Geckeler C, Lüthi M, et al. eProbe: sampling of environmental DNA within tree canopies with drones. *Environ Sci Technol.* 2024. <https://doi.org/10.1021/acs.est.4c05595>.
49. Tedersoo L, Bahram M, Zinger L, Nilsson RH, Kennedy PG, Yang T, et al. Best practices in metabarcoding of fungi: from experimental design to results. *Mol Ecol.* 2022;31:2769–95.
50. Groben G, Clarke BB, Kerkhof LJ, Bonos SA, Meyer WA, Qu Y, et al. Mycobiome analysis of tall fescue grass under drought stress using the Illumina MiSeq and Oxford Nanopore Technology MinION. *Phytobiomes J.* 2023;7:413–23.
51. Furneaux B, Bahram M, Rosling A, Yorou NS, Ryberg M. Long- and short-read metabarcoding technologies reveal similar spatiotemporal structures in fungal communities. *Mol Ecol Resour.* 2021;21:1833–49.
52. Stevens BM, Creed TB, Reardon CL, Manter DK. Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. *Sci Rep.* 2023;13:9323.
53. Mafune KK, Godfrey BJ, Vogt DJ, Vogt KA. A rapid approach to profiling diverse fungal communities using the MinION™ nanopore sequencer. *Biotechniques.* 2020;68:72–8.
54. Hu Y, Irinyi L, Hoang MTV, Eenjes T, Graetz A, Stone EA, et al. Inferring species compositions of complex fungal communities from long- and short-read sequence data. *MBio.* 2022;13:e02444.
55. Kudjordjie EN, Schmidt-Høier AS, Brøndum MB, Johnsen MG, Nicolaisen M, Vestergård M. Early assessment of fungal and oomycete pathogens in greenhouse irrigation water using Oxford nanopore amplicon sequencing. *PLoS ONE.* 2024;19:e0300381.
56. Mafune KK, Vogt DJ, Vogt KA, Cline EC, Godfrey BJ, Bunn RA, et al. Old-growth *Acer macrophyllum* trees host a unique suite of arbuscular mycorrhizal fungi and other root-associated fungal taxa in their canopy soil environment. *Mycologia.* 2023;115:470–83.
57. Taylor DL, Walters WA, Lennon NJ, Boichicchio J, Krohn A, Caporaso JG, et al. Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for Illumina amplicon sequencing. *Appl Environ Microbiol.* 2016;82:7217–26.
58. Edgar RC. UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat Methods.* 2013;10:996–8.
59. Edgar RC, Flyvbjerg H. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics.* 2015;31:3476–82.
60. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10.
61. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol Evol.* 2013;4:n/a–n/a.
62. Edgar RC. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *bioRxiv.* 2016;13:081257.
63. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019;37:852–7.
64. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics.* 2011;27:863–4.
65. Tedersoo L, Tooming-Klunderud A, Anslan S. PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytol.* 2018;217:1370–85.
66. Ni Y, Liu X, Simeneh ZM, Yang M, Li R. Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Comput Struct Biotechnol J.* 2023;21:2352–64.
67. Jain M, Tyson JR, Loose M, Ip CLC, Eccles DA, O'Grady J, et al. MinION analysis and reference consortium: phase 2 data release and analysis of R90 chemistry. *F1000Research.* 2017;6:760.
68. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* 2016;4:e2584.
69. White TJ, Bruns TD, Lee S, Taylor J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR—protocols and applications—a Laboratory Manual. Academic Press; 1990. p. 315–22.
70. Pölme S, Abarenkov K, Henrik Nilsson R, Lindahl BD, Clemmensen KE, Kauserud H, et al. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Divers.* 2020;105:1–16.
71. Nguyen NH, Song Z, Bates ST, Branco S, Tedersoo L, Menke J, et al. FUN-Guild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* 2016;20:241–8.
72. Tanuchai B, Ji L, Schroeter SA, Wahdan SFM, Hossen S, Delelegn Y, et al. FungalTraits vs. FUNGuild: comparison of ecological functional assignments of leaf- and needle-associated fungi across 12 temperate tree species. *Microb Ecol.* 2022;85:1–18.
73. R Foundation for Statistical Computing. A Language and Environment for Statistical Computing. 2022.
74. RStudio Team. RStudio: Integrated Development Environment for R. 2022.
75. Weiss S, Xu ZZ, Peddada S, Amir A, Bittiger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome.* 2017;5:27.
76. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
77. Mittelstrass J, Sperone FG, Horton MW. Using transects to disentangle the environmental drivers of plant-microbiome assembly. *Plant Cell Environ.* 2021. <https://doi.org/10.1111/pce.14190>.
78. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. *BMC Bioinform.* 2009;10:421.
79. Chao A, Chiu CH, Jost L. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through hill numbers. *Annu Rev Ecol Evol Syst.* 2014;45:297–324.
80. Jost L. Entropy and diversity. *Oikos.* 2006;113:363–75.
81. Li D. hillR: taxonomic, functional, and phylogenetic diversity and similarity through Hill Numbers. *J Open Source Softw.* 2018;3:1041.
82. Brooks M, Bolker B, Kristensen K, Maechler M, Magnusson A, Skaug H, et al. glmmTMB: Generalized Linear Mixed Models using Template Model Builder. CRAN: Contributed Packages. 2017.
83. Bergelson J, Mittelstrass J, Horton MW. Characterizing both bacteria and fungi improves understanding of the *Arabidopsis* root microbiome. *Sci Rep.* 2019;9:24.
84. Fox J, Weisberg S, Price B. car: Companion to Applied Regression. CRAN: Contributed Packages. 2019.
85. Lenth R V, Banfai B, Bolker B, Buerkner P, Giné-Vázquez I, Herve M, et al. emmeans: Estimated Marginal Means, aka Least-Squares Means. CRAN: Contributed Packages. 2017.
86. Deiner K, Walser JC, Mächler E, Altermatt F. Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biol Conserv.* 2015;183:53–63.
87. Crawley MJ, The R. The R Book. 2nd ed. Hoboken: John Wiley & Sons, Ltd; 2013.
88. Martino C, Morton JT, Marotz CA, Zengler K. A novel sparse compositional technique reveals microbial perturbations. *MSystems.* 2019;4:1–13.
89. Gloor GB, Fernandes AD, Macklaim JM, Albert A, Links M, Quinn TG, et al. Analysis of differential abundance taking sample and scale variation into account. *PLoS ONE.* 2022;8:60.
90. Kolde R. Package "pheatmap." 2022.
91. Dove S, Böhm M, Freeman R, Jellesmark S, Murrell DJ. A user-friendly guide to using distance measures to compare time series in ecology. *Ecol Evol.* 2023;13:e10520.
92. Cha S-H. Comprehensive survey on distance/similarity measures between probability density functions. *Int J Math Models Methods Appl Sci.* 2007;1:300–7.
93. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 2001;26:32–46.

94. Oksanen AJ, Blanchet FG, Kindt R, Legen- P, Minchin PR, Hara RBO, et al. Community ecology package. CRAN R-project. 2012. 263.
95. Martinez Arbizu P. pairwiseAdonis: Pairwise multilevel comparison using adonis. 2020.
96. Niskanen T, Lücking R, Dahlberg A, Gaya E, Suz LM, Mikryukov V, et al. Pushing the frontiers of biodiversity research: unveiling the global diversity, distribution, and conservation of fungi. *Ann Rev Environ Resour.* 2023;48:149–76.
97. Tedersoo L, Mikryukov V, Anslan S, Bahram M, Khalid AN, Corrales A, et al. The Global Soil Mycobiome consortium dataset for boosting fungal diversity research. *Fungal Divers.* 2021;111:573–88.
98. Ciuffreda L, Rodríguez-Pérez H, Flores C. Nanopore sequencing and its application to the study of microbial communities. *Comput Struct Biotechnol J.* 2021;19:1497–511.
99. Delahaye C, Nicolas J. Sequencing DNA with nanopores: troubles and biases. *PLoS ONE.* 2021;16(10):e0257521.
100. Ihrmark K, Bodeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, et al. New primers to amplify the fungal ITS2 region—evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol Ecol.* 2012;82:666–77.
101. Schlegel M, Queloz V, Sieber TN. The endophytic mycobiome of European ash and Sycamore maple leaves—geographic patterns, host specificity and influence of ash dieback. *Front Microbiol.* 2018;9:2345.
102. Toju H, Tanabe AS, Yamamoto S, Sato H. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE.* 2012;7:e40863.
103. Aguayo J, Husson C, Chancerel E, Fabreguettes O, Chandelier A, Fourrier-Jeandel C, et al. Combining permanent aerobiological networks and molecular analyses for large-scale surveillance of forest fungal pathogens: a proof-of-concept. *Plant Pathol.* 2020;70:181–94.
104. Moccia K, Papoulis S, Willems A, Marion Z, Fordyce JA, Lebeis SL. Using the microbiome amplification preference tool (MAPT) to reveal *Medicago sativa*—associated eukaryotic microbes. *Phytobiomes J.* 2020;4:340–50.
105. Viotti C, Chalot M, Kennedy PG, Maillard F, Santoni S, Blaudez D, et al. Primer pairs, PCR conditions, and peptide nucleic acid clamps affect fungal diversity assessment from plant root tissues. *Mycology.* 2024;15:255–71.
106. Oita S, Carey J, Kline I, Ibáñez A, Yang N, Hom EFY, et al. Methodological approaches frame insights into endophyte richness and community composition. *Microb Ecol.* 2021;82:21–34.
107. U'Ren JM, Lutzoni F, Miadlikowska J, Laetsch AD, Elizabeth Arnold A, Arnold AE. Host and geographic structure of endophytic and endolithic fungi at a continental scale. *Am J Bot.* 2012;99:898–914.
108. Bačková A, Cooke DEL, Milenković I, Májek T, Nagy Z, Corcobado T, et al. Hidden *Phytophthora* diversity unveiled in tree nurseries of the Czech Republic with traditional and metabarcoding techniques. *Eur J Plant Pathol.* 2024. <https://doi.org/10.1007/s10658-024-02886-1>.
109. Siddique AB, Khokon AM, Unterseher M. What do we learn from cultures in the omics age? High-throughput sequencing and cultivation of leaf-inhabiting endophytes from beech (*Fagus sylvatica* L.) revealed complementary community composition but similar correlations with local habitat conditions. *MycKeys.* 2017;20:1–16.
110. Carini P, Marsden PJ, Leff JW, Morgan EE, Strickland MS, Fierer N. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol.* 2017;2:16242.
111. Gamboa MA, Laureano S, Bayman P. Measuring diversity of endophytic fungi in leaf fragments: Does size matter? *Mycopathologia.* 2003;156:41–5.
112. Frey JE, Frey B, Frei D, Blaser S, Gueuning M, Bühlmann A. Next generation biosecurity: towards genome based identification to prevent spread of agronomic pests and pathogens using nanopore sequencing. *PLoS ONE.* 2022;17:e0270897.
113. Ohta A, Nishi K, Hirota K, Matsuo Y. Using nanopore sequencing to identify fungi from clinical samples with high phylogenetic resolution. *Sci Rep.* 2023;13:9785.
114. Linkies A, Graeber K, Knight C, Leubner-Metzger G. The evolution of seeds. *New Phytol.* 2010;186:817–31.
115. Stump SM, Sarmiento C, Zalamea PC, Dalling JW, Davis AS, Shaffer JP, et al. Colonization of seeds by soilborne fungi: linking seed dormancy-defense syndromes, evolutionary constraints, and fungal traits. In: Seed endophytes: biology and biotechnology. Springer International Publishing; 2019. pp. 465–81.
116. Franić I, Allan E, Prospero S, Adamson K, Attorre F, Auger-Rozenberg MA, et al. Climate, host and geography shape insect and fungal communities of trees. *Sci Rep.* 2023;13:1–13.
117. Christian N, Sedio BE, Florez-Buitrago X, Ramírez-Camejo LA, Rojas EI, Mejía LC, et al. Host affinity of endophytic fungi and the potential for reciprocal interactions involving host secondary chemistry. *Am J Bot.* 2020;107:219–28.
118. Schroeder JW, Martin JT, Angulo DF, Arias-Del Razo I, Barbosa JM, Perea R, et al. Host plant phylogeny and abundance predict root-associated fungal community composition and diversity of mutualists and pathogens. *J Ecol.* 2019;107:1557–66.
119. Pellitier PT, Zak DR, Salley SO. Environmental filtering structures fungal endophyte communities in tree bark. *Mol Ecol.* 2019;28:5188–98.
120. Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, et al. Phylogenomics and the rise of the angiosperms. *Nature.* 2024. <https://doi.org/10.1038/s41586-024-07324-0>.
121. Nilsson RH, Ryberg M, Wurzbacher C, Tedersoo L, Anslan S, Pölme S, et al. How, not if, is the question mycologists should be asking about DNA-based typification. *MycKeys.* 2023;96:143–57.
122. Tanunchai B, Ji L, Schroeter SA, Wahdan SFM, Thongsuk K, Hilke I, et al. Tree mycorrhizal type regulates leaf and needle microbial communities, affects microbial assembly and co-occurrence network patterns, and influences litter decomposition rates in temperate forest. *Front Plant Sci.* 2023;14:1239600.
123. Weiß M, Waller F, Zuccaro A, Selosse M-A. Sebaciales—one thousand and one interactions with land plants. *New Phytol.* 2016;211:20–40.
124. Hardoim PR, van Overbeek LS, Berg G, Pirttilä AM, Compant S, Campisano A, et al. The hidden world within plants: ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol Mol Biol Rev.* 2015;79:293–320.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.