Received: 28 September 2021 | Revised: 7 February 2022 | Accepted: 30 March 2022

DOI: 10.1111/mec.16460

INVITED REVIEW

MOLECULAR ECOLOGY WILEY

Best practices in metabarcoding of fungi: From experimental design to results

Leho Tedersoo[1,2] | Mohammad Bahram[1,3] | Lucie Zinger[4,5] | R. Henrik Nilsson[6] | Peter G. Kennedy[7] | Teng Yang[8] | Sten Anslan[9] | Vladimir Mikryukov[1,9]

[1]Mycology and Microbiology Center, University of Tartu, Tartu, Estonia

[2]College of Science, King Saud University, Riyadh, Saudi Arabia

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

[4]Institut de Biologie de l'ENS (IBENS), Département de Biologie, École normale supérieure, CNRS, INSERM, Université PSL, Paris, France

[5]Naturalis Biodiversity Center, Leiden, The Netherlands

[6]Department of Biological and Environmental Sciences, Gothenburg Global Biodiversity Centre, University of Gothenburg, Göteborg, Sweden

[7]Department of Plant and Microbial Biology, University of Minnesota, Saint Paul, Minnesota, USA

[8]State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing, China

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

Correspondence
Leho Tedersoo, Mycology and Microbiology Center, University of Tartu, 2 Liivi, Tartu 50409, Estonia.
Email: leho.tedersoo@ut.ee

Funding information
Horizon Europe EuropaBon project and Estonian Science Foundation, Grant/Award Number: PRG632

Handling Editor: Pierre Taberlet

Abstract

The development of high-throughput sequencing (HTS) technologies has greatly improved our capacity to identify fungi and unveil their ecological roles across a variety of ecosystems. Here we provide an overview of current best practices in metabarcoding analysis of fungal communities, from experimental design through molecular and computational analyses. By reanalysing published data sets, we demonstrate that operational taxonomic units (OTUs) outperform amplified sequence variants (ASVs) in recovering fungal diversity, a finding that is particularly evident for long markers. Additionally, analysis of the full-length ITS region allows more accurate taxonomic placement of fungi and other eukaryotes compared to the ITS2 subregion. Finally, we show that specific methods for compositional data analyses provide more reliable estimates of shifts in community structure. We conclude that metabarcoding analyses of fungi are especially promising for integrating fungi into the full microbiome and broader ecosystem functioning context, recovery of novel fungal lineages and ancient organisms as well as barcoding of old specimens including type material.

KEYWORDS
biodiversity, bioinformatics, community ecology, experiment setup, molecular identification, statistical analyses

1 | INTRODUCTION

Fungi have traditionally been identified based on macro- and micromorphological features of fruiting body specimens or pure cultures. The introduction of molecular techniques established in the late 1980s represented a significant leap forward in fungal identification. Especially PCR amplification combined with Sanger sequencing of the nuclear 18S (SSU) and 28S (LSU) ribosomal rRNA genes and the nuclear ribosomal internal transcribed spacer (ITS) region from fungal tissue (e.g., lichen thalli, lesions in plant and animal tissue, cultures from environmental samples and ectomycorrhizal root tips) quickly became popular and offered unprecedented taxonomic resolution.

Common uses included species- and genus-level identification, analysis of cryptic species, and phylogenetic assessment of major fungal clades as well as the kingdom Fungi at large (Gherbawy & Voigt, 2010). Later on, the identification of multiple fungi from more diverse substrates, including soil, plant roots and water became possible by including a cloning step of amplicons prior to sequencing. However, these studies usually operated with tens to low hundreds of reads, rarely numbering in the thousands required to properly estimate fungal diversity in soils (Taylor et al., 2014). Accordingly, sequences and operational taxonomic units (OTUs) were usually handled manually or using specific programs, with no urgent need for bioinformatics tools.

The development of high-throughput sequencing (HTS) methods such as 454 pyrosequencing (454 Inc., obsolete), Illumina sequencing (Illumina Inc., www.illumina.com) and Ion Torrent (Thermo Fisher Scientific Inc., www.thermofisher.com) transformed fungal identification capacity in the 2000s (Jumpponen & Jones, 2009). These so-called next- or second-generation HTS methods increased the sequencing capacity by 2–6 orders of magnitude and the number of simultaneously processable samples by 1–2 orders of magnitude. These metabarcoding methods (cf. Taberlet et al., 2012) enabled estimating fungal diversity exhaustively from environmental DNA (eDNA) on an individual sample scale as well as facilitating global-scale comparisons (Sun et al., 2021; Tedersoo et al., 2014). Yet, these second-generation platforms as well as the more recent DNA nanoball sequencing (DNBseq; MGI-Tech Inc., www.mgitech.com) were only able to sequence short (<550 bases) fragments of the genetic markers, resulting in the loss of taxonomic resolution and phylogenetic information as well as difficulties in identifying technical artefacts compared with longer Sanger reads.

In the 2010s, long-read, third-generation HTS platforms such as PacBio single-molecule real-time (SMRT) sequencing (Pacific BioSciences Inc., www.pacbio.com) and nanopore sequencing (Oxford Nanopore Technologies Inc., https://nanoporetech.com) were introduced (van Dijk et al., 2019). Due to their low sequencing depth (tens of thousands of reads in total, resulting in only hundreds rather than thousands of reads per sample) and high raw error rates (12–20%), these methods could not initially compete with short-read HTS platforms. However, both technologies made a great leap forward in 2020 when PacBio Sequel II instruments became broadly available and new solutions were developed to greatly reduce error rates in nanopore sequencing (Karst et al., 2021; Tedersoo et al., 2021). These long-read technologies and synthetic long reads provide high-quality sequence data for up to 3.5 kb amplicons, which enables bridging variable and conserved fragments of one or more genes in a single sequencing round as well as resolving alleles and haplotypes (Callahan et al., 2021; Tedersoo, Albertsen, et al., 2021).

Along with the rapid development of HTS methods, bioinformatic platforms and analytical resources evolved to match the computational needs imposed by large data sets. Metabarcoding approaches have been extensively reviewed in several recent studies with a focus on their conceptual foundation (Taberlet et al., 2018), pathogenic organisms (Piombo et al., 2021; Tedersoo et al., 2019), applications in mycology (Lindahl et al., 2013; Nilsson et al., 2018), eukaryotes more broadly (Ruppert et al., 2019) as well as overall experimental planning (Zinger, Bonin, et al., 2019), trade-offs among technology generations (Kennedy et al., 2018; Loit et al., 2019) and analytical pitfalls (Cristescu & Hebert, 2018; Halwachs et al., 2017). Here, we provide a review of available methods and propose best practices for designing and performing studies using metabarcoding in fungi. We also compare the performance of several popular methods developed for bacteria or animals to assess their suitability for fungi. The vast majority of our recommendations are relevant to prokaryotes, protists and metazoans alike.

2 | PLANNING A METABARCODING STUDY

To test scientific hypotheses, researchers should first consider a proper methodological experimental design – either observational, experimental or combined – including technical, analytical, personnel-related and financial requirements. Experimental designs of broad representativeness (e.g., geographical and ecological scope) and independence of replicates (i.e., no spatiotemporal autocorrelation) are strongly recommended (Gotelli & Ellison, 2013; Zinger, Bonin, et al., 2019). Indeed, metabarcoding studies do not differ from traditional ecological studies, in which the number and distribution of study sites must be defined appropriately depending on the initial question (Dickie et al., 2018). Additionally, metabarcoding studies require an optimal number of local, biological replicates that can be determined based on the variance reported in previous studies (Alteio et al., 2021) or pilot experiments. Intuitively, more replicates will be required when any expected ecological differences are relatively small or when the studied location exhibits strong spatial or environmental heterogeneity. Sometimes, pooling individual environmental samples is useful for studies covering larger geographic scales (Box 1).

The size of individual environmental samples should be large enough to secure enough material for DNA extraction and potential physicochemical analysis (e.g., pH and C/N ratio). It is also important to consider the amount of material from the perspectives of pretreatment and storage. Too much material will be difficult to mix, dry or freeze – and will prove costly to preserve in a buffer. To ensure statistical independence of samples within a site, samples should be located from outside the spatial autocorrelation range, which is typically 5–10 m in soil fungi (Bahram et al., 2013) and corresponds to the near-maximum size of most macrofungal individuals (Douhan et al., 2011). In aquatic habitats, communities are likely to compositionally autocorrelate for even larger distances (Matsuoka et al., 2019). When assessing diversity patterns along ecological gradients, transects (e.g., latitudinal, altitudinal and salinity gradients) should be replicated. Spatial independence should also be ascertained for plots and treatments. In field and laboratory experiments, this is best achieved by a randomized block design (Legendre & Legendre, 2012). A stratified block design may

BOX 1 Trade-offs in sample pooling

To improve representativeness of the samples at minimum extra cost, pooling statistically nonindependent subsamples is a widely used option. The number and spatial distance of subsamples may be of great importance to provide a representative view of the microbial diversity in heterogeneous habitats; less inclusive subsampling designs are likely to result in underestimating diversity (Figure Box 1). The number of subsamples to be pooled depends on the research question and the size of the area, with 7–25 being optimal in most cases (Schwarzenbach et al., 2007). Both physical and analytical pooling improve richness and composition assessments of soil fungi (Schwarzenbach et al., 2007; Song et al., 2015) and reduce estimated variance (Dickie et al., 2018). However, pooling of physical samples may result in the loss of patchily occurring rare taxa (e.g., in extremely dilute fish eDNA samples with a detection threshold of 0.05% of total relative abundance at deep sequencing; Sato et al., 2017). These results may be relevant for fungal groups of relatively low DNA content and/or rRNA copy numbers, for example, *Glomeromycota* and unicellular taxa. It is likely that the pooling effect depends on habitat heterogeneity, including pH, organic matter content, salinity and plant species present - all of which are factors known to affect fungal composition in different environments (Amend et al., 2019; Grossart et al., 2019; Nilsson et al., 2019; U'Ren et al., 2019). Therefore, pooling samples with potentially different microbial composition (e.g., leaves of different plant species) is not recommended. Theoretically, pooling does not work optimally in situations where the samples contain different amounts of DNA and where the low-DNA samples feature unique, rare species. Given the greater overall richness, pooled samples also require deeper sequencing to detect rare taxa. Furthermore, pooling is unsuited for co-occurrence analyses assessing biotic interactions (Bahram et al., 2014). Pooling individual samples at the site level (at the phase of DNA extraction, PCR, library preparation or sequence data) may be the most useful when these samples cannot be used as independent replicates (local- or landscape-scale spatial autocorrelation), for example, for regional- to global-scale analyses.

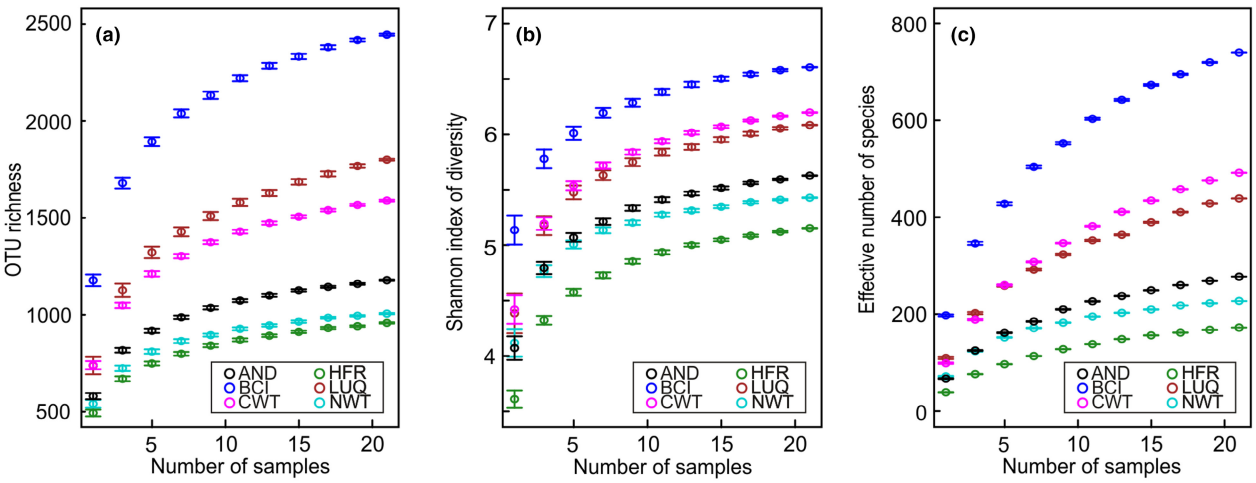

FIGURE BOX 1 Potential underestimation of biodiversity and high variance at low number of (sub)samples: (a) OTU richness, (b) Shannon index of diversity and (c) Effective number of species ($q = 1$; i.e., exponent of the Shannon index). Rarefied data sets (13,688 fungal ITS2 reads) from six sites (Zhou et al., 2016) were randomized 100 times to generate subsets of various sample sizes representing composite, analytically pooled samples. These samples were further rarefied at the same level for calculating diversity values. Note that the differences would be greater without rarefaction. Circles represent means of analytically pooled samples, and error bars indicate 95% CI. The differences between a single sample and 21 randomly selected samples average 104.1% (±20.7%), 36.2% (±5.4%) and 269.6% (±69.1%) for (a), (b) and (c), respectively. We conclude that pooling information from around 10 samples per site is sufficient to distinguish richness patterns, but 〉20 samples are needed to capture 〉90% of fungal richness based on this sampling scheme

be important in environments with known heterogeneity. It is advisable to collect samples in as short a time period as possible to avoid seasonal and weather effects such as freeze-thaw cycles and rainfall after a long dry period (for soil and leaves) that may cause rapid turnover of microbes and degradation of their DNA by moulds. All sampling locations (including positions in controlled experiments) and sampling dates should be recorded precisely to permit controlling for spatiotemporal effects in the following statistical analyses (Bahram et al., 2015; Tedersoo, Anslan, Bahram, Drenkhan, et al., 2020).

3 | SAMPLING AND STORING

We strongly recommend wearing disposable gloves during sampling to avoid contaminating samples with skin or forward microbiota. To reduce the risk of cross-contamination between independent samples in the field, sampling tools should be replaced or sterilized between sampling events with oxidizing agents (e.g., bleach; Fischer et al., 2016), DNase solutions or flame but not only with alcohols, as the latter do not denature DNA. Samples should be collected into clean containers such as paper bags (leaf and fruiting body material), plastic bags (roots, soil and sediments) or screw-cap vessels (soil, water and sediments). It is recommended to sample in the field during dry weather to avoid contamination by water from rain and wet gloves. To enable removal of site or sampling material contaminants a posteriori, it is also recommended to include field controls (e.g., empty tubes left opened at the site or extraction of sample storage buffer) in the experiment (Zinger, Bonin, et al., 2019). Finally, it is important to limit the biological activity within samples post-harvest (i.e., growth of fast-growing molds), which can be done by maintaining the samples at cold temperature during transport.

To obtain DNA of high quality, the best option is to either extract DNA right after sampling whenever possible or to rapidly freeze the collected materials in liquid nitrogen and maintain them at −80°C (U'Ren et al., 2014). Pooled subsamples should be well mixed before freezing, because it may subsequently be difficult to homogenize frozen material, which could lead to some parts of the sample or entire subsamples effectively being excluded from DNA extraction. When freezing, it is important to avoid thawing, which may lead to sample spoilage and significant changes in the detected communities (Anslan et al., 2021; Clasen et al., 2020). Long-term storage (2–4 weeks) at 4°C may alter soil fungal diversity (Delavaux et al., 2020) and promote proliferation of moulds (Clasen et al., 2020). Rapid drying methods such as freeze drying and cabinet drying are alternatives to freezing to prevent DNA degradation (Castaño et al., 2016). Drying with silica gel is a viable option for samples of a few grams (but see Guerrieri et al., 2021 for larger amounts) and will also work well for plant material such as thin leaves and fine roots. Importantly, liquid preservatives such as cetyltrimethylammonium bromide (CTAB), ethanol and specific DNA/RNA preservation solutions perform poorly for above-gram samples (e.g. Delavaux et al., 2020; Zaiko et al., 2022). Nevertheless, Longmire buffer (100 mm Tris, 100 mm ethylenediaminetetraacetic acid, 10 mm NaCl, 0.5% sodium dodecyl sulphate, 0.2% sodium azide; 1:1 vol/vol) works well for sediment and water samples, in which DNA otherwise would degrade very rapidly (Kumar et al., 2020). Dried material and samples fixed in buffers can be kept in the dark at room temperature. Samples can be stored long-term (decades) if kept air-tight in the dark and at constant temperature (Wang et al., 2021). It is also essential to store DNA samples, preferably in frozen or lyophilised form, for potential subsequent quality check, reanalyses for other research purposes or simply for a reanalysis using more sophisticated HTS methods in temporal studies (Jarman et al., 2018). Drying or lyophilisation is essential for pooled samples of coarse fragmented materials (e.g., wood chips and plant litter), which require grinding for adequate mixing.

4 | MOLECULAR ANALYSIS

4.1 | DNA extraction

Prior to DNA extraction, it is important to homogenise the material by using bead beating in microcentrifuge tubes, or a mortar and pestle or micropestles or knife mill for small sample numbers. The required amount of material should be weighted to the DNA extraction tube and the rest could be stored for backup or for example stable isotopes or chemistry analyses. It is usually undesirable to reach the full capacity of the DNA extraction kit, because several types of samples (e.g., peat soils, dead wood, debris-rich sediments and fleshy plant tissues) may absorb the liquid, or inhibitors may be co-extracted. For well-homogenised soil samples, there are only minor differences in perceived richness when using DNA extracts from 0.25, 1 or 10 g material (Song et al., 2015), but increasing the volume through replicate extractions or through more material using "maxi" kits provides more reproducible estimates (Dickie et al., 2018). It is crucial to perform weighing and DNA extraction under a dedicated laminar flow in a room separated from the PCR laboratory to avoid cross-contamination and air contamination by amplicons. Such potential contaminants can also be detected and removed in downstream analyses through analysis of blank DNA extraction controls.

For DNA extraction, we recommend following the protocols elaborated for relevant substrates, either manual methods or commercial kits. The CTAB and phenol-chloroform protocols (multiple variants exist) are the most broadly used manual methods for obtaining large quantities of long DNA molecules. While the quantity of DNA from the aforementioned protocols is usually relatively large, it is often less pure than kit-based approaches and so may require further dilution ahead of PCR to minimize the effect of inhibitors present in the sample (see below). Because of functional limitations in DNA extraction robots, the DNA purity and yield obtained with these protocols tend to be lower than with analogous nonrobot kits. As a rule of thumb, both commercial nonrobot and robot-based kits are roughly two and five times more time-efficient, but 2–10 times more costly compared with manual protocols.

Depending on the sample type and extraction method, the DNA may contain impurities that inhibit PCR amplification. These can be overcome by pretreatment of samples during DNA extraction (e.g., Al^{3+} or Ca^{2+} flocculation of humic substances; Braid et al., 2003), purification using specific kits (e.g., polyvinylpolypyrrolidone spin columns against humic and fulvic acids in soil; universal Zymo Research OneStep PCR Inhibitor Removal Kit or Macherey-Nagel NucleoSpin Inhibitor Removal Kit against polyphenolics, humic and fulvic acids, tannins and melanin) or equipment (e.g., SCODA electrophoresis), or precipitation with ethanol. Importantly, under most conditions, dilution of the DNA extracts may be sufficient to eliminate PCR inhibition (Wang et al., 2017). DNA concentrations can be increased

by precipitation with ethanol, salts (e.g., sodium acetate) and carriers (e.g., Pellet Paint Co-Precipitant, glycogen or linear polyacrylamide). If the DNA extract contains a large proportion of short fragments (e.g., degraded DNA due to poorly preserved samples or extracellular DNA) that hamper amplification and may promote chimera formation, these can be removed by elution using specific kits such as AMPure (Beckman Coulter Inc., www.beckman.com) and ProNex (Promega Corp., www.promega.com). Extracellular DNA may account for 80% of total DNA, but it has little effect on estimates of diversity, as it comes from active populations (Nagler et al., 2018) and the dead cells of indigenous biota (Lennon et al., 2018). However, it is crucial to remove extracellular DNA for time series and co-occurrence analyses on a microscale (Lennon et al., 2018), which can be performed by sample treatment with ethidium or propidium monoazide (Wagner et al., 2008). Substrates destined for mesh-bag experiments (plant litter, wood) can be relieved from unnecessary microbial DNA with gamma-irradiation (Brabcova et al., 2016) or dry heating at >120°C before exposure.

4.2 | Control samples

Negative controls for sampling, DNA extraction and PCR, and positive controls (including mock communities) improve scientific reproducibility by offering means by which to estimate the accuracy of the analyses (reviewed in Zinger, Bonin, et al., 2019). Negative and positive controls inform about external and cross-contamination as well as potential index-switching (Carlsen et al., 2012; Esling et al., 2015). Mock community analysis provides additional insights into the qualitative (i.e., estimation of PCR/sequencing error rates) and quantitative capacity (i.e., biased amplification) to recover the original diversity. Positive controls and mock communities may consist of artificial synthesised molecules or DNA extracts of actual species known not to occur in the experimental system (Ihrmark et al., 2012; Song et al., 2015). A sophisticated mock community should comprise >10 species with variable G+C content, amplicon length and quantity based on actual marker copy numbers. Additionally, due to index-switching issues, consideration of the specific species composition of the mock community is desirable. Specifically, if the mock community contains the same taxa present in the samples being analysed, it becomes impossible to determine whether any switched reads in the samples come from the mock community or not. One solution to this problem is to use a nonbiological mock community, containing multiple "species" that are synthetically constructed to have properties equivalent to biological species but that never are present in nature (Palmer et al., 2018). Mock community analyses commonly fail to recover all species and usually reveal more OTUs than the input because of variable DNA quality, PCR bias, trace contamination, index-switching and occasional divergent haplotypes (Bakker, 2018; Ihrmark et al., 2012). Therefore, failing to recover the initial mock community does not necessarily indicate that the analyses have failed, but it sheds light on potential biases and serves as a reference to correct the data through bioinformatics and/or post-bioinformatic processing.

4.3 | Genetic markers

Obtaining high-quality amplicons is one of the most important steps in metabarcoding analyses. This can be achieved by selecting a suitable genetic marker, polymerase, relevant primers and appropriate thermocycling conditions. Each PCR run requires a negative control to rapidly detect contamination.

The ITS region of the rRNA cistron is the most broadly used marker for fungi in both DNA barcoding and metabarcoding analyses due to its multiple copy numbers, optimal species-level resolution in most groups and the possibility to design both fungal and universal primers (Nilsson et al., 2018; Schoch et al., 2012). The ITS region is unsuited to target certain fungi such as Microsporidia (intracellular animal parasites) that may lack this region and certain Tulasnellaceae (orchid root symbionts) that have mutations in primer sites (Rammitsu et al., 2021; Tedersoo et al., 2015). Furthermore, ITS sequences lack variability in some species in certain species-rich genera comprising pathogens and saprotrophs such as *Trichoderma* and *Fusarium*, and their analysis requires using additional taxonomic markers, typically protein-coding genes (Cai & Druzhinina, 2021; O'Donnell et al., 2015). The arbuscular mycorrhizal *Glomeromycota* have multinucleate hyphae with highly variable ITS copies, which has rendered the rRNA 28S and 18S gene fragments of broad use as well (Kolaříková et al., 2021).

Because of sequencing read length limits imposed by second-generation HTS platforms, researchers have mostly focused on

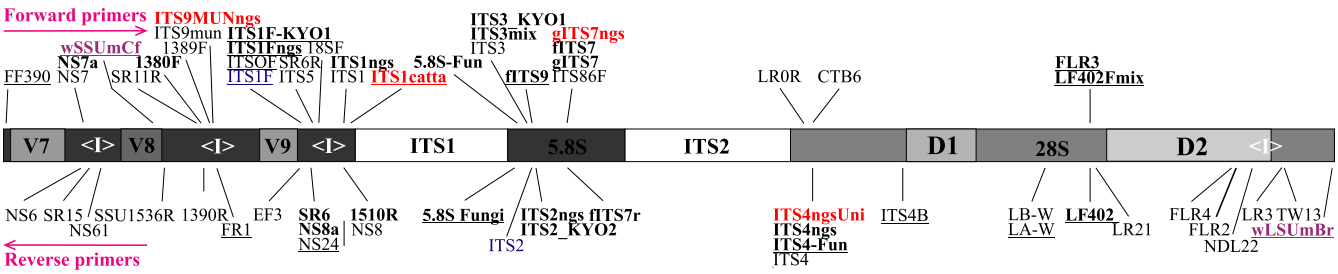

FIGURE 1 Primer map of the rRNA operon internal transcribed spacer (ITS) region. Primers developed for metabarcoding studies are in bold; fungus-specific primers are underlined. Primers used for global mycobiome projects are indicated in red (all fungi), purple (*Glomeromycota*) and blue (Earth Microbiome project), common intron sites. Updated from Nilsson et al. (2018)

either the ITS1 or ITS2 subregion, which taken separately have lower taxonomic resolution and do not offer as suitable primer sites as the full region (Tedersoo, Albertsen, et al., 2021; Tedersoo et al., 2015). Systematic differences in rRNA gene copy numbers and amplicon length among fungal taxa, due to introns in both ITS subregions, may bias the overall quantitative estimates of major fungal groups (Lofgren et al., 2019; Tedersoo et al., 2015).

For metabarcoding, ecologists use mostly primers designed decades ago for Sanger sequencing analyses (Figure 1; White et al., 1990). These original primers are not optimal for the many fungal groups that have one or more primer-template mismatches. They can be improved by adding degenerate positions to minimise primer bias (Tedersoo & Lindahl, 2016) and promote quantitative performance (Pinol et al., 2019). However, multiple degeneracies may require altering the 1:1 ratio of primers and may require extra PCR cycles, because not all primer variants match to templates. The broadly used fungus-specific forward primer ITS1F is particularly problematic because of several critical mismatches in certain groups of moulds and putative animal pathogens (Tedersoo & Lindahl, 2016). Researchers should also consider the common presence of an intron at the end of 18S rRNA gene, which prevents sequencing of the taxa with this intron (Figure 1). It may be important to pair primers with similar melting temperatures to attain optimal performance.

There are different amplicon library preparation strategies that require consideration during the primer design step (Figure 2). The metabarcoding primers may be equipped with both sample-specific index and platform-specific adapters for sequencing. The alternative strategy is to use shorter primers with only sample-specific indexes, which are c. 30–40% cheaper and easier to amplify but require specific library preparation depending on the sequencing platform. Approaches requiring several PCR steps are also available (Figure 2; Bohmann et al., 2022), but these are more prone to contamination and chimera formation. Although vulnerable to contamination, the use of combinations of Illumina flow cell indices in the second PCR step enables ultra-high multiplexing of samples without index-switching bias (Holm et al., 2020).

The sample-specific indexes are typically 6–14 bases in length and differ from each other by at least four nucleotides (including indels) for error correction (Buschmann & Bystrykh, 2013). Their GC content should be in the range of 25–75% and homopolymers >2 nucleotides should be avoided. An example of >300 indexes is listed in Taberlet et al. (2018). To reduce amplification biases, there should be a 2–3-base linker between the index and PCR primer, which should not align to any of the targeted sequences. The quality of Illumina sequencing benefits from heterogeneity spacers added to the indexes (Figure 2; Fadrosh et al., 2014). To secure more equal library preparation, indexes should start with the same nucleotide. The same indexes (but not linkers) can be used with multiple primers, but each primer-index combination should be tested for hairpin structure formation in silico. Indexing both primers with unique tags (i.e., unique dual indexing) is more expensive, but allows users to greatly reduce index-switching artefacts (Schnell et al., 2015), and is therefore strongly recommended.

4.4 | Polymerases

With respect to DNA polymerases, proofreading polymerases have much-reduced error rates and therefore result in fewer spurious OTUs (Bakker, 2018; Oliver et al., 2015). The 3′ to 5′ exonuclease activity of proofreading polymerases performs primer editing in the last 4–6 nucleotide positions, reducing primer bias (Gohl et al., 2021). However, this activity varies by polymerase, the mismatching nucleotide and probably concentration of inhibitors (Gohl et al., 2021), and the effect on multiple near-terminal mismatches remains unexplored. Hence, proofreading polymerases may also strongly reduce the specificity of taxon-specific primers. Furthermore, the exonuclease activity of proof-reading polymerases creates multiple short fragments, especially at low ddNTP concentration and prolongs elongation times, which may result in more chimeras already at early stages of the PCR process (Ahn et al., 2012). For longer amplicons, it is crucial to select high-fidelity polymerases to secure amplification completion and hence reduce production of chimeric artefacts (Heeger et al., 2018). Thus, a wise selection of primers and polymerases allows researchers to obtain the same amount of high-quality data with lower sequencing depth.

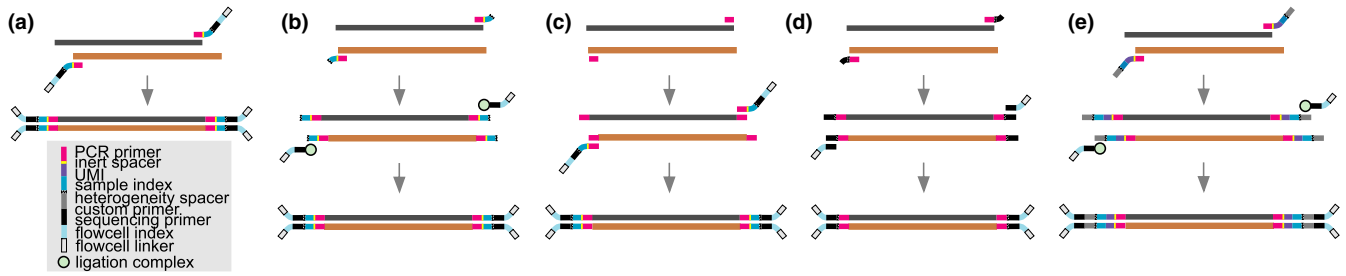

FIGURE 2 Common and perspective library preparation strategies for Illumina sequencing: (a) adding indexes and adapters using amplification with fusion primers, (b) adding indexes with amplification and adapters by ligation, (c) amplification and then adding indexes and adapters in second PCR step, (d) indexing samples with combinations of Illumina indices (Holm et al., 2020), and (e) incorporating unique molecular identifiers (UMIs) in the first amplification step (modification of Karst et al., 2021). Libraries of other HTS platforms require more specific protocols

4.5 | Thermal cycling conditions

Regarding PCR conditions, reducing annealing temperature may promote amplification of targeted taxa that have one or more primer-template mismatches, but it may also enhance nonspecific priming, resulting in amplification of random genomic fragments or untargeted taxa. The number of PCR cycles should be kept at below 30 – optimally resulting in a weak band on a gel – to minimize formation of chimeric molecules and distorting relative abundances (D'Amore et al., 2016; Gohl et al., 2016; Lindahl et al., 2013). Rather than losing samples with no visible amplicons, it is advised to add a few extra cycles to problematic samples, but users should keep in mind that these low-input or inhibitor-rich samples have elevated risk of contamination or biased diversity patterns (Eisenhofer et al., 2019). Adding bovine serum albumin (BSA) may be useful for improving amplification success but distort the retrieved community (Zaiko et al., 2022). Thus, the impacts of newly designed specific primers, polymerases, blocking oligonucleotides (cf. Vestheim et al., 2011), stabilisers and altered thermal cycling conditions on community composition should be carefully tested to avoid potential biases.

Biological samples may differ several orders of magnitude in their DNA content, quality and abundance of inhibitors. For PCR, the DNA content is rarely equalised, because typically 80–99% of eDNA is nonfungal, and the fungal fraction may vary significantly across samples (Bahram et al., 2018; Tedersoo et al., 2015). There is no consensus on whether or how the DNA quantity should be standardised, although diluted samples may yield a higher proportion of contaminants (Lindahl et al., 2013) as well as relatively lower diversity and greater variability (Castle et al., 2018, but see Song et al., 2015 and Wang et al., 2017). Therefore, at least two PCR replicates are needed to account for the stochasticity. Such technical replicates can be pooled for further analysis steps (Alberdi et al., 2018; Lindahl et al., 2013), but this pooling step will prevent evaluation of the PCR replication and exclusion of dysfunctional amplicons (Taberlet et al., 2018).

4.6 | Alternatives to traditional eDNA amplicon-based methods

To focus on the active community, RNA instead of DNA can be used as a target for sequencing (Singer et al., 2017, but see Blazewicz et al., 2013 for limitations). One option is to amplify reverse transcribed cDNA, which is also applicable for ITS sequences in spite of the short life of precursor RNA (Rajala et al., 2011). Interestingly, cDNA-based HTS reveals multiple taxa not recovered using DNA and vice versa (Rajala et al., 2011). Another option is direct RNA sequencing, which is currently provided only by Oxford Nanopore Technology (ONT; Garalde et al., 2018). Both methods produce more errors than state-of-the-art DNA-based methods. As both PacBio and ONT sequencing make it possible to record modified nucleotides with various methylations, it may be possible to record various artificial nucleotide analogues (e.g., 3-bromo-deoxyuridine) incorporated into DNA in real time (Georgieva et al., 2020; Hanson et al., 2008). Stable isotope probing (SIP) is widely used for bacteria because of their rapid metabolism of ^{13}C-enriched substrates (Berry & Loy, 2018), but it has been little used in mycology (but see Hannula et al., 2017; Lopez-Mondejar et al., 2020). RNA-based SIP applications may offer more promise in fungi than for bacteria (Ghori et al., 2015; Singer et al., 2017).

Metagenomics and metatranscriptomics can be used for large-scale identification of organisms. These methods are free from PCR biases but may be affected by library preparation biases and add an order of magnitude to the costs (Quince et al., 2017; Singer et al., 2017). While these methods work reasonably well on bacteria and viruses with small and densely packed genomes and for which a rich set of reference genomes are available, analyses of fungi and other eukaryotes are heavily biased because of highly different genome sizes, number of rRNA gene copies and the striking lack of reference genomes for many important groups (Geisen et al., 2015; Tedersoo et al., 2015). This may change very soon by activities of the ongoing Earth Biogenome project (Lewin et al., 2022) and the use of taxonomically more informative long reads for reference (Simmons et al., 2020). Furthermore, DNA/RNA quantity combined with metagenomic or metatranscriptomic information enables to quantify the biomass and potential functionality of different microbial groups (Bahram et al., 2018; Žifčáková et al., 2016). Alternatively, a spike-in method can be used for biomass estimates. This method relies on adding artificial molecules or cells (known amount) of one or preferably more species (absent from biological samples) before DNA extraction (Palmer et al., 2018; Rao et al., 2021), and quantification based on the proportions of reads.

4.7 | DNA library preparation

Among-sample variability of amplicon quantity is high at a low number of PCR cycles. Therefore, the amounts of amplicons should be standardised for improved comparability of sequencing depths. This is most efficiently achieved by DNA capture on a solid phase with limited binding capacity (SequalPrep, Thermo Fisher Technologies; Harris et al., 2010).

The equimolarly pooled samples are subjected to library preparation using HTS platform-specific kits. Aside from multiple kits for Illumina, those free from amplification steps and biases of G+C content and fragment length are recommended (Bowers et al., 2015; Sato et al., 2019). Amplicons produced by different primers, even when of similar length, should not be mixed into the same library because of great differences in yield (Tedersoo et al., 2015), but it will add to library preparation costs. In-house library preparation may be up to 5-fold cheaper compared to commercial services. For small laboratories, it is advisable to order the library preparation service from a sequencing company to avoid contamination and reduce the risk of sequencing failure.

4.8 | Sequencing platforms

For metabarcoding, both the second-generation and third-generation platforms can be considered. Currently, the second-generation platforms allow sequencing up to c. 550-base markers, but their throughput exceeds that of third-generation platforms by 1–2 orders of magnitude and their costs per base are at least an order of magnitude lower. Given their relative accuracy, Illumina (HiSeq and NovaSeq instruments in 2 × 250 paired-end mode and MiSeq) and MGI-Tech (DNBSEQ-G400RS in 2 × 200 paired-end mode) are best suited for analyses of short barcodes such as ITS1, ITS2 or one or two variable regions combined within the 18S and 28S rRNA genes (Tedersoo, Albertsen, et al., 2021).

The average raw read length of PacBio and ONT instruments exceeds 20 kb. The libraries of PacBio consist of circularised amplicons, which are sequenced multiple times (circular consensus sequencing; CCS) such that error rates decrease from 10–15% to <0.1% at >10-fold consensus. This allows high-quality sequencing of up to 3.5 kb fragments that cover multiple rRNA markers. Such long reads offer much improved taxonomic resolution and allow rigorous phylogenetic analyses based on reasonably long alignments of conserved regions (Simmons et al., 2020). Furthermore, random PCR and sequencing errors are typically ironed out during the clustering process (Tedersoo et al., 2018), and much of the relatively more degraded extracellular DNA is excluded.

Currently, ONT sequencing does not offer sufficient read quality for metabarcoding. Although unique molecular identifiers (UMIs) can be used in the generation of consensus sequences (Figure 2e; Karst et al., 2021), obtaining at least 20-fold consensus will reduce throughput and increase the overall cost tremendously. UMIs can also be used for producing synthetic long reads using any of the short-read platforms, which results in principally error-free long reads (Callahan et al., 2021). However, a new commercial LoopSeq service provided by Loop Genomics, Inc. (www.loopgenomics.com) is relatively costly (43–100 USD/sample) and requires validation for eukaryote studies. Taken together, the choice of HTS strategy depends on expected data quality, number of samples included, desired sequencing depth and amplicon length as well as available financial resources (Tedersoo, Albertsen, et al., 2021).

5 | BIOINFORMATIC DATA ANALYSIS

5.1 | Quality-filtering

The raw output of sequencing instruments is converted to the fastq format, which is compatible with all major quality-filtering tools. As most bioinformatic platforms have been developed for bacterial 16S data, these differ greatly in their capacity to handle fungal ITS sequences, which typically cannot be reliably aligned much beyond the genus level (Anslan et al., 2018). Based on citations, the most broadly used platforms include QIIME2, mothur, PIPITS, SEED2, SCATA and PipeCraft. Features as well as pros and cons of the most popular and recently developed platforms are presented in Table 1.

For Illumina and MGI-Tech instruments that produce paired-end reads, it is recommended to assemble these paired reads, unless the amplicon is longer than the paired reads combined (e.g., Bissett et al., 2016). Generally, it is advisable to disregard unpaired reads because of the risk of low read quality in their distal end.

A universal step of quality-filtering includes demultiplexing, trimming primer and index sequences, and removal of low-quality and nontarget reads. Any ambiguous nucleotides and mismatches to indexes or primers are indicative of potentially low read quality and these reads could be excluded. Dual-indexed reads with mismatching pairs or a missing index from one primer are indicative of index switching and compromised sequence data, respectively. The size of full-length ITS ranges from 250 (some *Saccharomycetales*) to around 1500 bases (e.g., some *Cantharellales* and various unicellular groups), but *Microsporidea* may have only a few bases of rudimentary ITS sequences. The ITS1 and ITS2 subregions taken separately vary from 50 to around 1000 bases. There is also great length variation in 18S and 28S rRNA genes, which is mostly ascribed to introns.

Some 16S-based workflows recommend removal of homopolymers >6 bases (e.g., default in QIIME2), but the ITS region of many fungal and other eukaryotic taxa commonly harbour homopolymers exceeding 10 bases. Hence, homopolymer length must not be used as an indicator in quality-filtering for ITS sequences and other noncoding regions.

For ITS metabarcoding, it is important to remove the flanking 18S and 28S rRNA genes, because these conserved ends display no species-level resolution and random errors in these regions complicate clustering (Lindahl et al., 2013). Furthermore, chimeric breakpoints may be common in these regions but are nearly impossible to recognize from such short, 10–70-base fragments. ITS extraction can be performed using ITSx (Bengtsson-Palme et al., 2013) or ITSxpress (Rivers et al., 2018) that cut out ITS1, ITS2 and full-length ITS region based on kingdom-wide Hidden Markov Models (HMMs); ITSxpress is several times faster than ITSx (Rivers et al., 2018).

Chimeric molecules are mainly generated in the excessive cycles of PCR and are therefore nearly always less abundant than their parent molecules (Sze & Schloss, 2019). They are usually represented by singletons and doubletons restricted to a single sample (Tedersoo, Mikryukov, et al., 2021). There are multiple algorithms for chimera detection, of which UCHIME (Edgar et al., 2011) is by far the most universal and widely used. It is recommended to perform chimera filtering both in de novo and reference-based modes, which compare OTUs against each other (in ranked abundance within a sample) and against a reference database (e.g., UNITE), respectively. According to our experience, a vast majority of reference-based chimeras are true chimeras, whereas c. half of the de novo chimeras may be false positives (Aas et al., 2017; Tedersoo, Mikryukov, et al., 2021). Not all chimeras are detected by the programs, so it may be advisable to remove all singletons or OTUs with <5 or <10 reads in the case of deep sequencing (Edgar, 2016). Chimeric molecules usually have only a partial match to the reference sequence (coverage 55–98%)

TABLE 1 Properties of bioinformatic pipelines used for fungal metabarcoding

Software	Interface	Type	Target group	ITS extraction	Unique features	References
QIIME2	CL, graphical (alpha release)	Offline, multiple included dependencies	Prokaryote, short-read	ITSxpress	Interactive data visualisation; Deblur filtering	Bolyen et al. (2019)
Mothur	CL	Offline	Prokaryote, short-read	–	OptiClust clustering	Schloss (2020)
Claident	CL	Offline	Universal, short-read	ITSx	Noise removal options	Tanabe and Toju (2013)
PIPITS	CL	Offline	Fungi, ITS1, ITS2	ITSx	–	Gweon et al. (2015)
SEED2	Graphical	Offline	Universal, short-read	ITSx compatible	Read visualisation	Větrovský et al. (2018)
PEMA	CL	Offline, multiple included dependencies	Universal, short-read	–	Parameter file for automated pipeline	Zafeiropoulos et al. (2020)
PipeCraft	Graphical	Offline, multiple included dependencies	Universal, short/long-read	ITSx	Multiple HTS platform- and data-specific options	Anslan et al. (2017)
DADA2	CL (R-Studio)	Offline	Universal, short/long-read	–	ASV calculation	Callahan et al. (2016)
LotuS2	CL	Offline	Universal, short/long-read	ITSx	Low RAM usage, false positives removal	Ozkurt et al. (2021)
Galaxy	Graphical	Online	Prokaryote, short-read	–	Links to other Galaxy functions	Jalili et al. (2020)
FROGS	Graphical	Online	Prokaryote, short-read	ITSx	Galaxy functions	Escudie et al. (2018)
AMPtk	CL	Offline	Universal, short-read	–	Hybrid method for taxonomy assignment; spike-in control	Palmer et al. (2018)
OBITools	CL	Offline	Universal, short-read	–	–	Boyer et al. (2016)
SCATA	Graphical	Online	Fungi, short/long-read	ITSx	Unique clustering options	https://scata.mykopat.slu.se/
gDAT	Graphical	Offline	Universal, short-read	–	–	Vasar et al. (2021)
Cascabel	CL	Offline	Universal, short-read	–	Graphical report	Abdala Asbun et al. (2020)
MICCA	CL	Offline	Universal, short-read	–	OTUCLUST clustering algorithm	Albanese et al. (2015)
Nf-core/ampliseq	CL	Offline	Universal, short-read	ITSx	Easy parallelisation, reproducibility	Straub et al. (2020)
CloVR-ITS	CL	Offline/online	Fungi, short-read	–	Virtual machine-based with the possibility to run on Amazon EC2 Cloud	White et al. (2013)
DAnIEL	Graphical	Online	Fungi, short-read	ITSx (via PIPITS)	Raw data download from SRA	Loos et al. (2021)

Abbreviation: CL, command line.

at high sequence similarity (>90%), which allows additional manual detection of rare, potentially artefactual OTUs.

Index-switches (also known as tag-switches, index-jumping and index cross-talk) are the most deleterious phenomenon in HTS data, because they result in technical cross-contamination among samples and may blur especially patterns in host specificity, taxon networks and biogeographical patterns (Calderon-Sanou et al., 2020; Carlsen et al., 2012). Index switches occur during PCR, T4 blunt-ending and

BOX 2 Exact sequence variant (ESV) methods

ESV approaches represent a specific type of greedy de novo clustering and rigorous removal of noise and rare haplotypes (Callahan et al., 2017) to calculate essentially 100%-similarity OTUs. Several alternative methods - notably DADA2, UNOISE and deblur – have recently become popular in microbiology including fungal ecology (Glassman & Martiny, 2018). In DADA2, sample-wise rare variants are assigned to dominant haplotypes based on a Poisson error model for quality score and nucleotide transition model combination (Callahan et al., 2016). UNOISE performs a similar process less stringently by ignoring quality scores (Edgar, 2016). Deblur uses sequencing error profiles and is relatively less conservative than the other methods (Amir et al., 2017), but it requires equal length of amplicons and cannot be therefore used for ITS marker-based analysis. The ESV approaches are certainly useful for separating as many species/haplotypes as possible based on conserved genes, but their utility for ITS and protein-coding genes is unclear (Antich et al., 2021). They may outperform traditional OTU clustering approaches in distinguishing very closely related species of *Ascomycota* with haploid genomes. However, an ESV approach severely biased species richness estimates of metazoans based on the cytochrome oxidase 1 (CO1) gene (Antich et al., 2021; Brandt et al., 2021), and it is expected to perform poorly for fungal groups with dikaryotic (*Basidiomycota*), diploid (most unicellular groups) or polyploid (*Glomeromycota*) genomes that commonly exhibit two or multiple different rRNA gene and ITS copies per genome or even within haploid nuclei (Egan et al., 2018; Lindner et al., 2013; Runnel et al., 2022). Estensmo et al. (2021) demonstrated that in polypores, single species contained multiple ESVs. By reanalysing a data set from Furneaux et al. (2021), we show that the DADA2 ITS pipeline and UNOISE ESV approaches reduce phylogenetic richness by disproportionately eliminating rare members of the unicellular fungal groups, *Glomeromycota* and nonfungal eukaryotes (Figure Box 2). In terms of community composition, the results are similar between ESV and OTU-based approaches (Glassman & Martiny, 2018; Porter & Hajibabaei, 2020), because these are driven by abundant taxa. We conclude that ESV approaches overestimate richness of common fungal species (due to haplotype variation) but underestimate richness of rare species (by removing rare variants; see also Joos et al. 2020). ESV approaches can nevertheless be useful for studying allele or haplotype distribution of various common species based on eDNA (Zizka et al., 2020).

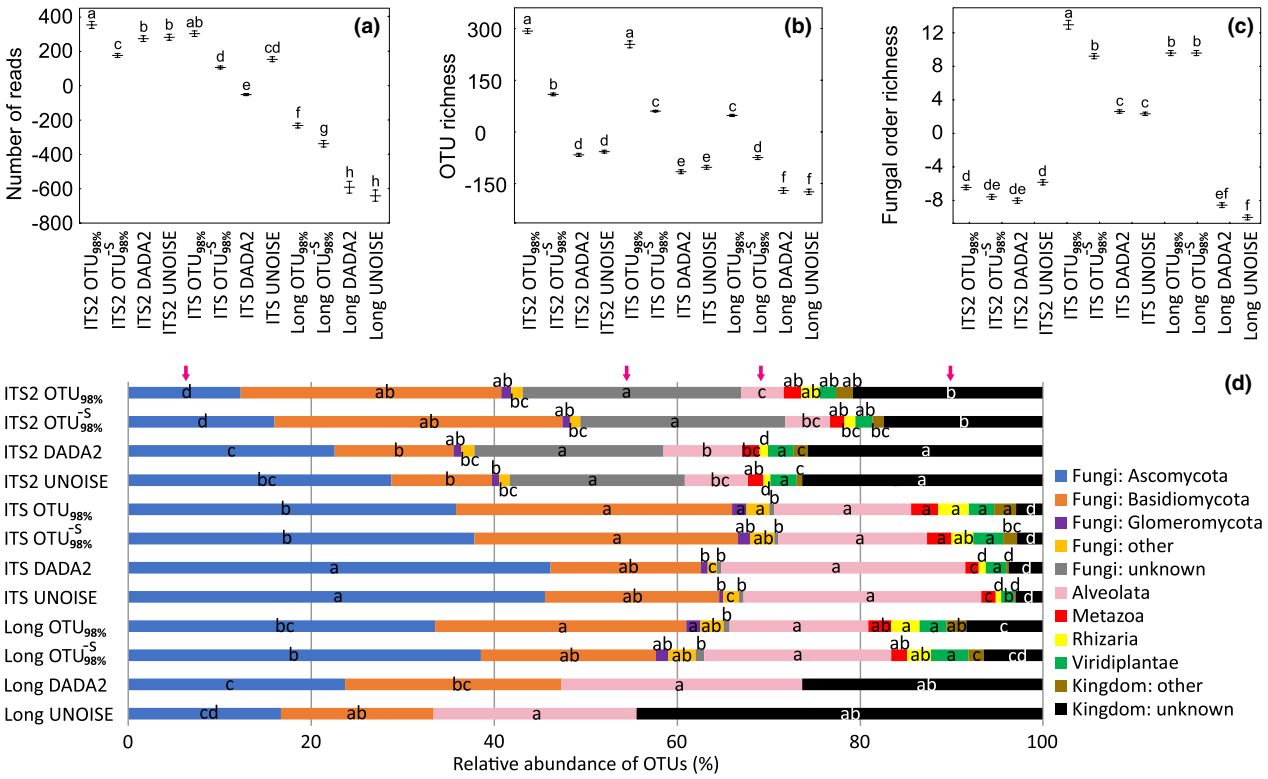

FIGURE BOX 2 Comparison of traditional OTU-based (OTU^{-s}, singletons removed) and ESV-based approaches (DADA2 and UNOISE) for inferring taxonomic richness using the ITS2 subregion, full ITS region and ITS +28S based on the data set of Furneaux et al. (2021): (a) average number of reads retained; (b) richness of OTUs; (c) richness of fungal orders; and (d) kingdom and phylum-level distribution of OTUs across the entire data set. In (d) letters indicate statistically significant differences among groups based on Scheffe PostHoc tests following sample-wise testing of log-ratio transformed proportions using three-way ANOVAs including sample (random blocking factor), marker length and analysis type: OTUs, OTUs without singletons (both: VSEARCH v.2.17.0, 98% sequence similarity threshold, following Tedersoo, Mikryukov, et al., 2021), DADA2 ESVs (v.1.20.0; default settings) and UNOISE3 ESVs (implemented in LotuS2 v.2.18; default settings). Arrows indicate groups with greatest differences. We conclude that: (1) the ESV approaches recover lower proportions of non-Dikarya and nonfungal taxa compared with traditional approaches; (2) analyses based on the ITS2 region alone revealed a higher proportion of OTUs that could not be identified to a fungal phylum or eukaryote kingdom compared with longer marker fragments due to the lower amount of sequence data available for taxonomic assessment; and (3) longer markers had fewer sequences passing the quality control and revealed relatively more singletons, suggesting accumulation of artefacts across the entire 1500 base amplicon

cross-pairing of amplicons from different libraries, and they are known from all sequencing platforms (Carøe & Bohmann, 2020; Schnell et al., 2015). Careful indexing of samples ameliorates this issue, but roughly 0.01–0.1% of obvious switches will remain. Index-switches can be assessed with an ad hoc score using the UNCROSS algorithm (Edgar, 2018), the unspread python script (Larsson et al., 2018), by tracking nonbiological spike-ins (Palmer et al., 2018) or by a positive control sample. Based on the distribution of spike-ins or positive control in biological samples and vice versa, index-switch rates can be estimated. Sequence abundances below the index-switching threshold are converted to zero. Index-switches and other contamination should be checked for each sequencing library separately. It is also useful to estimate occurrence of taxa in pseudosamples represented by unused indexes (see Taberlet et al., 2018; Zinger et al., 2021) to assess the proportion of index-switch-like artefacts resulting from mutations or primer impurities.

5.2 | Clustering

Clustering is used for aggregating reads into OTUs based on user-defined sequence similarity thresholds, usually ranging from 95% to 100% sequence similarity. Clustering methods differ by the means how sequences are aggregated into OTUs (reviewed in Lindahl et al., 2013; Taberlet et al., 2018), with single-linkage (reads grouped when any pairwise comparisons exceed the sequence similarity threshold), complete linkage (reads grouped when all pairwise comparisons exceed the threshold), average linkage (compromise between these methods) and greedy clustering algorithms (membership linked to distance from OTU centroid) recognised. Many algorithms are based on a global alignment or expect equal read length, and are thus unsuited for ITS data. Most algorithms enforce strict sequence similarity and coverage thresholds, but some allow relaxed overall similarity (e.g., swarm; Mahe et al., 2022) and do not consider end-gaps (indels in the end; e.g., SCATA, https://scata.mykopat.slu.se). Some algorithms allow truncation of homopolymers (e.g., the algorithms implemented in SCATA and PipeCraft), which may be warranted for the PacBio, ONT and Ion Torrent platforms where indels in homopolymers are the most common sequencing errors. Notably, current version of the relaxed-similarity swarm algorithm performs differently for taxa with relatively short or long ITS markers, which makes it unsuited for analyses of fungi at large, but it warrants testing for the capacity to distinguish amongst sister species and analytical errors at sequence similarities >98%.

Three types of clustering are commonly seen in metabarcoding efforts: open-reference, closed-reference and de novo clustering. In de novo clustering, the sequence data are clustered within the project. For closed-reference clustering and open-reference clustering, a reference database is used, but only in the latter method, unique sequences are retained. Open-reference, abundance-based greedy clustering or average clustering algorithms are recommended for generating the most stable OTUs by accounting for the taxonomic structure in the reference data (Cline et al., 2017; He et al., 2015).

Following studies in bacterial ecology and Sanger sequencing-based mycorrhizal ecology, a sequence similarity threshold of 97% is most commonly used, especially for short reads. Other studies have found that 98% or 98.5% sequence similarity for greedy clustering and single-linkage methods is a better compromise between sequencing errors and biological differences among closely related species and intraspecific variation (Kõljalg et al., 2013; Kyaschenko et al., 2017; Tedersoo et al., 2014). Yet, many closely related species of ascomycete saprotrophs and pathogens differ by no or only a few bases in full-length ITS (e.g., O'Donnell et al., 2015). Based on 18S rRNA gene sequences, ascomycete and basidiomycete species belonging to different orders may display identical or near-identical sequences, which render SSU unsuited for fungal DNA metabarcoding at species level. However, unicellular fungal lineages and *Glomeromycota* display greater SSU variation, although species-level distinction in the latter group is not straightforward as most virtual taxa and morphospecies are >50 million years old (Bruns & Taylor, 2016; Perez-Lamarque et al., 2020), which roughly corresponds to the age of genera in most other fungal groups. To improve taxonomic

resolution and reproducibility of identified taxa, exact sequence variant (ESV; also referred to as amplicon sequence variant) approaches have been developed for inferring OTUs (Box 2).

5.3 | Taxonomic assignments

A representative sequence is chosen from each OTU for taxonomic annotation. Most programs select one of the longest sequences, the consensus sequence or the most common sequence type for comparison to the reference corpus. In most cases, the latter is biologically the most meaningful. The longest sequence may contain artefactual insertions, untrimmed fragments of flanking genes or represent an unrecognised chimera. Consensus sequences are prone to lending voice to rare and perhaps compromised sequence data.

There are different approaches for taxonomic assignments, all of which require a well-curated reference database. The most common approach is custom BLAST searches (Camacho et al., 2009), where all representative sequences are compared pairwise to sequences in the reference database. Users can specify BLAST parameters from slow and stringent to rapid and discontinuous alignment. We recommend using word size <10 to be able to obtain long query-to-template alignments and hence the most precise estimates of e-value and sequence similarity. Alternative to BLAST, programs performing k-mer search such as Naïve Bayesian classifier (Porras-Alfaro et al., 2014; Wang et al., 2007) and SINTAX (Edgar, 2016) are up to 100-fold faster. These methods and the alignment-based PROTAX-Fungi (Abarenkov et al., 2018) offer probabilistic estimates for taxonomic precision. Kraken2 retrieves the lowest common ancestor for metabarcoding and metagenomics data based on multiple best hits (Wood et al., 2019). These algorithms work well in situations where reference data are abundant and accurately identified to species hypothesis or genus level, which is a somewhat atypical situation in the fungal kingdom (Lücking et al., 2021). Phylogenetic placement algorithms such as EPA-ng (Barbera et al., 2019) map rRNA gene-based OTUs to pre-established phylogenies, but these methods are not suited for the ITS region due to its hypervariability in both length and composition, and issues with computational speed. Some pipelines (e.g., AMPtk; Palmer et al., 2018) return taxonomic assessments based on results from multiple algorithms. While often slightly more conservative, this approach gives greater confidence in the assigned taxonomy when there is clear congruence across different algorithms. If the samples include many undescribed/unbarcoded species, we recommend relying on BLAST search against the UNITE database, focusing on the 5–10 best hits. Based on multiple global data sets, we have developed recommended taxon-specific e-values and sequence similarity thresholds for 18S-V9, ITS2 and full-ITS reads at the level of genera to phyla (Tedersoo et al., 2014; Tedersoo, Mikryukov, et al., 2021; updated in Table S1), but there are no automated ways of assigning these to OTUs. An important topic for future development is recognition of pseudogenes and nonfunctional rRNA gene variants in reference sequence databases and HTS data sets to reduce the number of artefactual OTUs (Porter & Hajibabaei, 2021).

With respect to reference databases, UNITE is the largest by containing curated data obtained from the International Nucleotide Sequence Databases consortium (INSDc) as well as data submitted directly to UNITE (Nilsson et al., 2018). Furthermore, UNITE provides species hypotheses (SHs) for ITS-based OTUs of fungi and other eukaryotes to enable unambiguous doi-based cross-communication of taxa among studies and across time (Kõljalg et al., 2016, 2020). Another curated data set is the Warcup training set, which covers a smaller set of well-identified fungi of mostly plant-associated *Ascomycota* and *Basidiomycota* (Deshpande et al., 2016). We recommend identification of fungal and other eukaryote ITS sequences based on the UNITE reference data set, because it is the largest curated database and it includes multiple nonfungal reads to facilitate separation of fungi from other eukaryotes (Anslan et al., 2018). We recommend the SILVA database for 18S and 28S rRNA gene reads (Quast et al., 2013), although their eukaryote taxonomy is outdated. For *Glomeromycota* 18S and 28S rRNA gene reads, MaarjAM (Öpik et al., 2010) and the AM-LSU pipeline (Delavaux et al., 2021) can be used, respectively.

5.4 | Functional assignments

Functional assignments can be supplied directly to best-hitting reference sequences and species, genera and orders based on their identification. The FUNGuild database is largely focused on the genus level (occasionally on the species level) and provides functional assignment of lifestyle and life mode based on probabilistic estimates (Nguyen et al., 2016). The FungalTraits database is focused on genus and order-level functional estimates covering additional traits such as fruiting body and hymenium type and capacities of performing certain biotic functions (Põlme et al., 2020). FungalTraits also allows complementary trait estimates (geographic distribution, isolation source and mycorrhizal type) based on sequence accessions and SHs. The Fun^{Fun} database includes a large number of genomic and enzymatic functional traits, but its taxonomic coverage is limited (Zanne et al., 2020). None of these databases should be taken as the golden truth, because infraspecific and infrageneric variation in fungal ecological traits are still poorly characterised.

From a functional perspective, we argue that it makes the most sense to analyse diversity at the level of all fungi and major functional guilds such as ectomycorrhizal, arbuscular mycorrhizal and putatively plant pathogenic fungi. Phyla, classes and orders are of limited value, because functionality of fungi is mainly conserved at the level of genus and family (Zanne et al., 2020). However, there are many examples where saprotrophic species are congeneric with ectomycorrhizal (e.g., *Meliniomyces* and *Entoloma*) or pathogenic (e.g., *Gibberella* and *Dothiostroma*) species. Yet, particularly orders and phyla are generally well-delimited and informative for understanding the phylogenetic structure. One notable challenge with functional assignments becomes apparent when individual taxa are matched to multiple ecological lifestyles (e.g., both saprotroph and endophyte). Lumping these taxa into a "multi-guild" category for ecological

analyses is not an effective solution; instead, it is recommended that users carefully analyse the associated primary literature for those taxa and determine, which functional type is most probably represented based on the study system or research question being addressed. This can be semiautomated using the PlutoF (Abarenkov et al., 2010) and Global Biotic Interactions (GloBI) databases (Poelen et al., 2014) or text mining software (Compson et al., 2018).

5.5 | Curation of the OTU matrix

The LULU software can be used for assessing co-occurrences of closely related taxa in the OTU-by-sample matrix (Frøslev et al., 2017). This program allows the removal of OTUs that potentially represent minority haplotypes of common OTUs and any remaining OTU-specific PCR and sequencing errors.

The OTU-by-sample matrix alone or tagged with sample metadata and taxonomic and functional annotation requires some manual curation in a spreadsheet program. Although many functions can be performed by Python and Bash scripting, the process of checking taxonomic annotations based on multiple best hits needs to be performed manually because the first hit may be incompletely annotated and curated reference databases still contain some misassigned data. Often, representative sequences require BLAST-based reanalysis against INSDc for taxonomic determination, more complete functional annotation (Fernandez et al., 2017) or additional chimera control using the graphic summary view (Nilsson et al., 2012). Manual BLAST examination of the 25–50 largest (most frequent or abundant) OTUs is a good way to identify compromised sequence data and to refine taxonomic annotation of OTUs.

Checking the distribution of sequences in control samples and spill-over of positive control in experimental samples is mostly manual work, but relevant functions are implemented in the metabaR R package and its associated tool vignette (Zinger et al., 2021). These steps are required to estimate the rate of index-switching and undertake appropriate measures. OTUs found in negative control samples should be assessed carefully, because these may be derived from molecular reagents, laboratory space or neighbouring samples (Eisenhofer et al., 2019; Loit et al., 2019). There are several programs for removing/subtracting contamination-affected OTUs from data cells (e.g. McKnight et al., 2019; Zinger et al., 2021). However, among-sample cross-contamination not affecting control samples (or if there are no controls) may be very difficult to find. Ranking OTUs by abundance and inspecting Spearman correlation matrices may be useful to detect such cross-contamination. To ensure recognition of contaminants, sample-specific spike-in molecules inserted to extractable samples can be used and traced to reads (Lagerborg et al., 2021).

Rarefaction is a commonly used option to standardise an OTU matrix to equal sequencing depth based on random subsampling of reads. Samples are typically rarefied to the lowest sequencing depth (after removal of failed or notably low-abundance samples), but there is no consensus whether the sequencing depth should reflect all reads, fungal reads or reads of each functional group taken separately. The main issue with rarefaction is the substantial loss of data that often corresponds to 90% of the sequencing depth, which may carry important information about diversity (McMurdie & Holmes, 2014). Rarefaction, especially to lower sequencing depths, provides relatively lower statistical power in the multivariate analyses as well (Figure 3; Martino et al., 2019). Another possibility is rarefaction to median sequencing depth (de Carcer et al., 2011) or model data using algorithms such as edgeR-TMM (Weiss et al., 2017). Another alternative to rarefying is scaling with ranked subsampling (SRS), which retains around 20% more OTUs (Beule & Karlovsky, 2020). Instead of rarefaction, we recommend accounting for the sequencing depth in models. However, some technical methodological comparisons and noncovariate models may still require rarefying.

6 | STATISTICAL DATA ANALYSES

6.1 | Alpha diversity metrics

Most mycologists as well as plant and animal ecologists strive to approximate the species level in biodiversity analyses, because macroscopic members of these groups can be well distinguished by morphology (Antich et al., 2021; Stajich et al., 2009). This contrasts to bacteria, where species are much more difficult to define. Molecular diversity of organisms has been traditionally studied using OTU richness as well as Shannon and Simpson indices and the related effective species numbers (Hill numbers; Alberdi & Gilbert, 2019; Chao et al., 2014). Diversity indices down-weight the effect of rare OTUs and therefore only weakly correlate with sequencing depth. OTU richness has a cumulating function with sequencing depth that is particularly prominent in diverse, pooled samples. Unless rarefaction is performed, it is important to include square-root or log-transformed sequencing depth (whichever is more informative) as a covariate. In large data sets, log-transformation of fungal OTU richness accounting for sequencing depth yields better-explained statistical models compared with untransformed data, rarefied data and diversity indices (Tedersoo et al., 2022). We do not recommend OTU richness extrapolations (e.g., Chao1 and ACE) even though they are commonly used, because they rely on the number of the rarest OTUs, which are commonly artefactual (Balint et al., 2016; Bunge et al., 2014).

Adding phylogenetic information to taxonomic composition eliminates the uncertainty regarding OTU calculations (Washburne et al., 2018) and reduces the effect of any remaining PCR/sequencing errors in the data (Taberlet et al., 2018). The ITS region is not amenable to robust multiple alignments and phylogenetic reconstruction much beyond the genus level. Therefore, phylogenetic measures require inferring phylogenetic distance matrices that may rely on ultrametric trees of conserved gene(s) (Davison et al., 2015; Horn et al., 2014), grafting phylogenies (Fouquier et al., 2016) or mapping of OTUs to distance-weighted phylogenies (Perez-Izquierdo et al., 2019) or hierarchical taxonomic trees (Chalmandrier et al., 2019; Tedersoo, Sánchez-Ramírez, et al., 2018). We recommend analysis of the ITS region and flanking, phylogenetically informative 18S or 28S rRNA genes for species-level

identification and phylogenetic placement, respectively. Shifts in phylogenetic diversity can be studied using standardised phylogenetic diversity (PD; averaged unique branch length), mean phylogenetic distance (MPD), UniFrac distance and mean nearest taxon distance (MNTD), the latter emphasising genus-level similarities. Testing phylogenetic conservatism, overdispersion and turnover across phylogenetic scales (Tucker et al., 2017) may be informative in analyses of plant-fungal interactions (Chalmandrier et al., 2019), fungal community assembly processes (Roy et al., 2019) and phylogeographic patterns (Turon et al., 2020). These measures and similar indices for community turnover can be calculated in phylocom (Webb et al., 2008), the R packages picante (Kembel et al., 2010), S.phylomaker (Qian & Jin, 2016) and PhyloMeasures (Tsirogiannis & Sandel, 2016) as well as other open-access scripts (Chalmandrier et al., 2019).

6.2 | Statistical methods

HTS analyses produce semiquantitative abundance data for OTUs (Amend et al., 2010), and many rare taxa remain below the detection level (Song et al., 2015). Fungal species differ greatly in the number of cells per unit biomass, the number of nuclei per cell and the

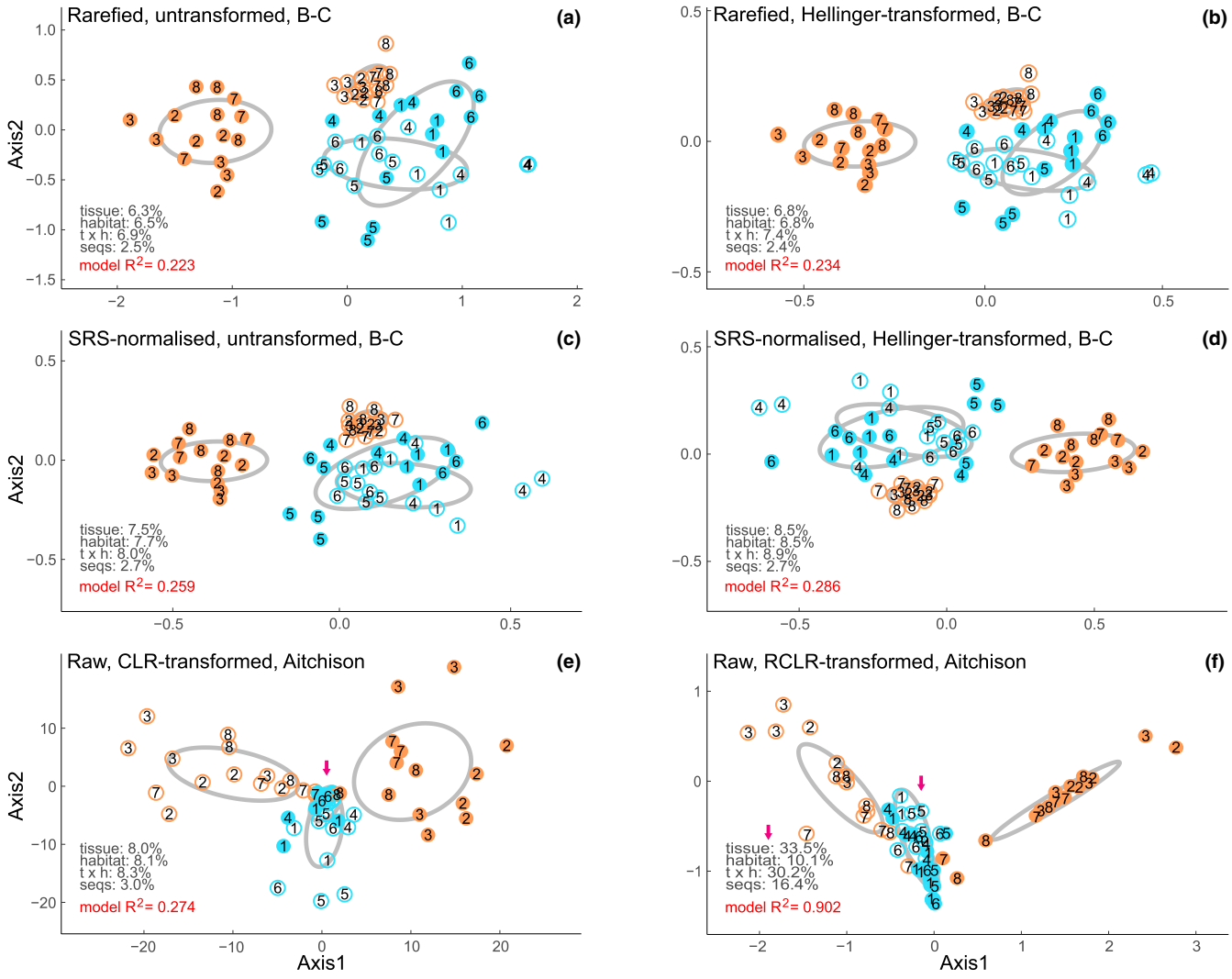

FIGURE 3 Nonmetric multidimensional scaling (NMDS) graphs illustrating relative performance of various normalisation methods and dissimilarity (B-C, Bray-Curtis; or Aitchison) measures in recovering trends in microbial eukaryote composition using untransformed and Hellinger-transformed data matrices in plant roots (filled circles) and leaves (open circles) in terrestrial (orange) and aquatic (blue) habitats: (a, b) rarefied data; (c, d) scaling with ranked subsampling (SRS) normalised data; (e) centred log-ratio (CLR) transformed data; and (f) robust CLR (RCLR) transformed data. Numbers on symbols indicate plant species (separate numbering for terrestrial and aquatic plants); ellipses depict 95% CI around means for tissue and habitat combinations. Explained variation (%) as revealed from Permanova+analysis is indicated in the bottom of each panel (t × h, tissue and habitat interaction; seqs, sequencing depth). Plant species effects are not analysed here for simplicity. Arrows highlight critical differences for interpretation. Data from A. Azadnia, V. Mikryukov, L. Tedersoo (unpublished data) that assessed the effect of aquatic versus terrestrial habitat and root versus leaf tissue on endophytic fungal diversity using eight plant species ($n = 4$) and PacBio sequencing of the full-length ITS region. We conclude that (1) rarefying underestimates treatment effects, and (2) RCLR transformation magnifies treatment effects by efficiently accounting for data compositionality and sparsity

number of ITS copies per nucleus that all affect relative read numbers (Lofgren et al., 2019). The use of quantitative methods assumes that all samples are influenced by these biases in a similar manner. Since there is a huge variety of methods, we provide an overview of alternative, suitable approaches rather than proposing single best practices.

For statistical analyses, most of the useful methods that have been elaborated in plant and microbial ecology are applicable for fungi, too (for review, see Balint et al., 2016; Buttigieg & Ramette, 2014; Hugerth & Andersson, 2017). In particular, the R packages vegan (Oksanen et al., 2019), phyloseq (McMurdie & Holmes, 2013) and microeco (Liu et al., 2021) are useful for routine analyses besides commercial statistical platforms. It should be kept in mind that in any ecological tests, the most biologically informative information to the readers is the effect size (either Z-score, ω^2 or adjusted R^2) and the direction and shape of the relationship. Values of statistical tests and *p*-values are context-dependent, albeit necessary to report along with variance and sample size for assessing the validity of the analysis. The rise of type I error rate in multiple testing needs to be accounted for by using flexible correction methods such as Benjamini-Hochberg correction (Noble, 2009).

Phylogenetic relationships among associated plants, metazoans or other organisms can be accounted for by testing phylogenetic signals or explicitly quantifying the phylogeny effect by using eigenvectors (e.g., adespatial package of R; Dray et al., 2018). Compared with using pure phylogenetic distances, the use of eigenvectors has proved to extract larger proportions of plant effects on fungal communities (Tedersoo et al., 2013; Yang et al., 2019). Similar eigenvector approaches have also been used to quantify and account for spatial as well as temporal and plant phylogeny effects on fungal diversity (Tedersoo, Anslan, Bahram, Drenkhan, et al., 2020; Zimmerman & Vitousek, 2012).

Metabarcoding analyses produce large amounts of data that require appropriate transformation. Classical transformations of metadata such as logarithmic (concentrations), square-root (counts) and log-ratio transformations (proportions) are necessary to shift the distribution of residuals towards normal distribution and reduce heteroscedasticity – principal assumptions of most parametric tests (Legendre & Legendre, 2012). Log-ratio transformation (LRT) also increases the independence of measurements of various proportions summing up 100% (see below).

When it comes to large amounts of metadata with potential multicollinearity, a rule of thumb is to remove variables with correlation coefficient >|0.7| to a stronger, more meaningful explanatory variable. Model selection can be performed by machine learning algorithms (Qu et al., 2019) such as random forest implemented in randomforest (Liaw & Wiener, 2002) and VSURF (Genuer et al., 2019) and/or R packages with step-wise model selection or the AICc information criterion (e.g., nlme R package; Pinheiro et al., 2011). The party package of R allows estimating interactions among two or more variables using machine learning (Strobl et al., 2009), which can be used to assess conditional and synergistic effects among variables (Rillig et al., 2019). For variables expected to exhibit unimodal (e.g., pH over a broad gradient) or cumulative (e.g., host richness and rainfall) relationships, inclusion of second-order polynomials or fitting generalized additive models (GAMs; mgcv package of R; Pedersen et al., 2019) may be more appropriate. Univariate analyses can also be performed using generalised linear modelling (GLM) by selecting a link function that best matches the distribution of variables (Leite & Kuramae, 2020).

For community-level analyses, it is important to consider that the sequencing data are compositional - that is, sequence counts represent at best relative rather than absolute biological abundances of organisms (Gloor et al., 2017; Lin & Peddada, 2020b). Therefore, it may be necessary to use Aitchison distance metric (Martino et al., 2019; Quinn et al., 2019) and LRT, of which robust centred LRT (RCLR) accounts for data sparsity constraint and differences in sequencing depth (Martino et al., 2019). The corresponding program DEICODE (https://github.com/biocore/DEICODE) is implemented in QIIME2 and PipeCraft2. With appropriate transformations and accounting for sequencing depth, different methods reveal roughly comparable results (Figure 3). Alternatively, programs for compositional data analyses (e.g., ANCOM-BC; Lin & Peddada, 2020a) can be used, but these have very limited options. PERMANOVA is currently the state-of-the-art multivariate analysis method; it outperforms common ordination methods because of its explicit hypothesis testing and allowing hierarchical design and among-variable interactions (Anderson, 2001). The program PERMANOVA+included in the Primer6/Primer7 packages (Anderson et al., 2008) offers more functionalities than the adonis routine of the vegan package. The rdacca.hp R package also estimates the overall importance of single predictors in multivariate models (Lai et al., 2022). Nonmetric multidimensional scaling (NMDS), principal coordinates analysis (PCoA) and redundancy analysis (RDA) are among the most popular ordination methods for two-dimensional visualisation of multivariate patterns (Paliy & Shankar, 2016). In particular, the ordiellipse function in the vegan package highlights within-group variance. General dissimilarity modelling (GDM) facilitates testing of nonlinear effects of multiple variables, which helps users to identify and understand critical biological thresholds (gdm R package; Manion et al., 2018). Generalised linear latent vector models (GLLVMs) also allow testing nonlinear responses by using variable link functions and including quadratic terms and among-species interactions in the models (Leite & Kuramae, 2020). Output of the gllvm R package provides residual environmental effects on individual OTUs and their correlations (Niku et al., 2019). For taxon-level analyses such as species (OTU) distribution modelling, random forest and univariate models are appropriate, but nonparametric tests should be used because of multiple zero-values. In addition, indicator species analyses can be used as implemented in the indicspecies R package (Caceres & Legendre, 2009). Specialist and generalist features of OTUs can be tested in two community types using CLAM, which outperforms other indicator statistics (Chazdon

et al., 2011). However, these indicator analyses may be sensitive to data compositionality.

Network analysis has become increasingly popular in microbial ecology (Faust, 2021; Mikryukov et al., 2021). Unipartite networks (co-occurrence analyses) are often used to infer positive (mutualism and facilitation) and negative (avoidance and competition) interactions among co-occurring taxa (Weiss et al., 2016). Bipartite networks estimate partner specificity and how these specific and nonspecific taxa are distributed (R package bipartite; Dormann et al., 2009). Both approaches also provide information on the association structure – modularity, nestedness and connectivity – and enable detection of network hubs, putative keystone taxa, higher-order interactions and association with environmental variables (Faust, 2021). However, these methods are sensitive to data structure and compositionality, and there is a great risk of overinterpretation (Blanchet et al., 2020; Matchado et al., 2021; Rao et al., 2021; Weiss et al., 2016). The main issue with comparing networks is related to their lack of true replication (Bahram et al., 2014) and dependence of results on the linkage metric, filtering threshold and network construction algorithm (Connor et al., 2017; Weiss et al., 2016). Nonetheless, consistent associations and integral topological parameters usually remain unaffected (Röttjers & Faust, 2018; Toju et al., 2015). For addressing biologically meaningful facilitation and avoidance strategies among species, the samples should be unpooled and of relevant size to ensure direct contact among organisms. Submillimeter scale is the most relevant for assessing fungal interactions with other fungi and bacteria in an abiotic environment. To reduce the number of false positive associations, it is recommended to reduce the matrix size (i.e., exclude rare taxa) to reach a fill level of c. 50% (Weiss et al., 2016). There are many network construction algorithms (Matchado et al., 2021; Weiss et al., 2016) for estimating relationship strength with correlation measures, indices of dissimilarity between species pairs, proportionality, or measures of conditional dependence omitting indirect connections or constructing consensus networks based on different measures (e.g., MENA, Deng et al., 2012; and CoNet, Faust & Raes, 2016). Some methods account for data compositionality and sparsity, for example SparCC (Friedman & Alm, 2011), SPIEC-EASI (Kurtz et al., 2015) and SPRING (Yoon et al., 2019). For network visualisation, Cytoscape (Shannon et al., 2003) and Gephi (Bastian et al., 2009) offer various options.

Structural equation modelling (SEM), in particular path analysis, tests the directionality as well as direct and indirect effects among variables (Collier, 2020; Fan et al., 2016). These are important to consider when the explanatory variables affect each other (e.g., vegetation and soil) or there are several related response variables (Delgado-Baquerizo et al., 2016; Yang et al., 2017). Nonetheless, the causal relationship identified by SEM models strongly relies on the hypothetical causalities tested, which should hence be properly justified by other empirical observations or theoretical foundations. In addition, SEM models have several commonly ignored assumptions: multivariate normality, linear associations, no missing data, no multicollinearity and large sample size – at least 20 samples per variable in the model (Collier, 2020). Taxonomic composition can also be included as principal components in SEM (Antoninka et al., 2009). The program AMOS (www.ibm.com) and the R package piecewiseSEM (Lefcheck, 2016) offer most functionalities needed for such analyses.

6.3 | Visualisation of results

There are a large number of methods and software tools available for visualising the statistical results. In all types of illustrations, it is important to emphasize the main differences (e.g., by colour, but using palettes suitable for various colour-blindness forms) with appropriate variance measures (if applicable) and avoid too much extra detail (i.e., noise). Network diagrams and circos plots (chord diagrams; Gu et al., 2014) are useful for displaying the structure of associations in the data.

Among simple graphs, box plots and scatter plots are the most commonly used. Violin plots are a specific type of box plots that indicate distribution of measurements and deviation from normality. Rarefaction curves (smoothed species accumulation curves) are useful for graphical comparisons of species evenness and richness but also evaluating sufficiency of sampling and sequencing depth within and among samples (Colwell et al., 2004). Heat maps are useful for visualising large correlation matrices or the results of multiple multifactorial analyses (e.g., ClustVis web tool; Metsalu & Vilo, 2015), although the typical lack of within-treatment variation can make their ecological interpretation challenging. Venn diagrams are useful for showing unique and shared variation among factors or OTUs across factor levels or combinations (e.g., venny; Oliveros, 2007), and rank-abundance plots can provide useful information on what specific taxa underlie certain ecological patterns. Similarly, overlying environmental vectors on ordination plots can help to identify the abiotic and biotic variables that are associated with specific taxon or sample abundances. In the R environment, the ggplot2 (Wickham, 2011) and phyloseq packages provide broad opportunities for constructing high-quality graphs.

The overall taxonomic composition is best visualised in stacked plots, which makes it easier to display multiple treatments. However, as error bars are lacking (there is no space), such plots and heat maps are examples of implicit pseudoreplication (but see figure in Box 2). Krona charts (Ondov et al., 2011) and heat trees (metacoder package of R; Foster et al., 2017) provide an efficient way of demonstrating the distribution of dominant taxa by taxonomic ranks (e.g., Nilsson et al., 2017). Because of high space requirements, the overall view or comparison of up to two levels of a treatment can be effectively indicated for Krona charts, but interactive versions can be provided as supplementary items. Heat trees can handle two factor levels or a single gradient.

While phylogenetic trees are generally too large for visualising taxonomic affiliation of OTUs in the main article, these are well suited to supplementary items. Although family-level and higher-level phylogenetic relationships cannot always be assessed based

on ITS data, these are suitable for demonstrating rough phylogenetic placement. Phylogenetic trees make more sense for 18S and 28S rRNA gene data and are furthermore helpful in detecting artefactual OTUs based on ultra-long branches or branches with zero length next to a long-branch taxon (Tedersoo et al., 2020). Large circular phylogenies squeezed into c. 17 cm page width offer limited opportunities for interpretation if taxon names are unreadable and branch support values are lacking. Using iTOL (interactive tree of life; https://itol.embl.de/), more than 50,000 taxa in phylogenetic trees can be tagged with large amounts of metadata for display in supplementary materials (Letunic & Bork, 2021).

6.4 | Data management

Funding requests should be written to cover not only the field and laboratory parts, but also data processing, metadata annotation and public deposition and dissemination. Fungal metabarcoding data should be submitted to any public archive such as short read archive (SRA), European nucleotide archive (ENA) or UNITE, indicating database accession numbers. UNITE calculates 100%-similarity OTUs for full-length ITS sequences for each biological sample; these are incorporated in the SH system and used for further reference along with the metadata. Direct submission of representative HTS sequences to INSDc is discouraged, because these poorly annotated and commonly low-quality data hamper further analyses. For data sharing, OTU community matrices, metadata and demultiplexing information should be uploaded in public repositories (e.g., Figshare, Dryad, zenodo) or supplementary materials, which enable other users to perform meta-analyses and populate databases (Põlme et al., 2020; Větrovský et al., 2020). For the sake of clarity and machine-readability, it is best to use standardised MIMARKS and MIxS terminology (Yilmaz et al., 2011). Scripts used for analyses should also be released in, for example, Github or zenodo, to secure reproducibility and potential reuse in other applications. We recommend researchers to use specific workflow managers to ease these procedures (Wratten et al., 2021).

In the materials and methods section of scientific articles, it is important to document all aspects of the analysis (Lindahl et al., 2013). This is likely to shorten the review process and help reviewers and readers to evaluate the validity and novelty of the procedures. For knowledge exchange, we endorse posting preprints at the submission stage to a preprint server such as bioRxiv or Authorea. Publishing open access may increase readership and citations several-fold since some universities in developing countries lack subscriptions to many journals (Piwowar et al., 2018).

7 | FUTURE PERSPECTIVES AND CONCLUSIONS

HTS analyses have recently opened new frontiers in many fields of mycology. Metabarcoding analyses using SIP of labelled substrates (Hannula et al., 2017) or coupled with metatranscriptomics (Žifčáková et al., 2016) have revealed functionally active fungi and their activity in situ. Parallel studies of fungi, bacteria and protists have shed light into antagonistic interactions (Bahram et al., 2018; Bork et al., 2015), the structure of the micro- and mycobiome web (Tipton et al., 2018) and mechanisms of community assembly (Zinger, Taberlet, et al., 2019). HTS-derived data have revealed several groups of previously undescribed (or unsequenced) order- and class-level fungal lineages (Tedersoo, Anslan, Bahram, Kõljalg, et al., 2020; Zhang et al., 2021). HTS reads offer material for constructing taxon-specific primers and probes for visualising cells (Chambouvet et al., 2019) and help discovery and characterisation of poorly known fungal lineages. Increasing read length and accuracy of HTS methods enhance taxonomic precision and highlight a venue for population-level studies based on eDNA (Byrne et al., 2017; Turon et al., 2020). This may be of particular relevance to fungal taxa that are difficult to culture and form no fruit-bodies (Lücking et al., 2021). Metabarcoding of short fragments furthermore sheds light into ancient DNA including potentially ancient fungi (Balint et al., 2018; Talas et al., 2021). Similarly, metabarcoding approaches are useful in generating DNA barcodes from organisms that are polyploid (Maeda et al., 2018) or harbour multiple haplotypes (Runnel et al., 2022), or are represented by century-old specimens including valuable type material where extra care is needed to identify and dismiss air-borne contaminants (Forin et al., 2018).

Through short-read metabarcoding, fungal diversity has been analysed in nearly all habitats on Earth, including extreme environments (Nilsson et al., 2018). We hope that with the assistance of growing reference databases, studies involving fungal taxonomic, phylogenetic and functional composition will flourish, because many principal aspects in evolutionary and functional (e.g., trait-based) ecology remain poorly known so far. These studies should be supplemented by rigorous experiments to validate the findings and infer causality. It somehow seems unacceptable that studies on one of the most important groups of nutrient cyclers – the fungi – be based on anything but the best and most up-to-date methodological recommendations, and we sincerely hope that this review has contributed to that end.

ACKNOWLEDGEMENTS

We thank Ning Daliang, Jizhong Zhou and Brendan Furneaux for sharing data and Liis Tiirmann for editing the manuscript. We are indebted to three anonymous referees who helped us to improve the manuscript substantially. The core funding is provided from the Horizon Europe EuropaBon project and Estonian Science Foundation (grant PRG632).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Leho Tedersoo, Mohammad Bahram, Sten Anslan and Vladimir Mikryukov reanalysed data. All authors contributed to ideas and writing.

DATA AVAILABILITY STATEMENT

This review produced no data and no code.

ORCID

Leho Tedersoo https://orcid.org/0000-0002-1635-1249
Mohammad Bahram https://orcid.org/0000-0002-9539-3307
Sten Anslan https://orcid.org/0000-0002-2299-454X
Vladimir Mikryukov https://orcid.org/0000-0003-2786-2690

REFERENCES

Aas, A. B., Davey, M. L., & Kauserud, H. (2017). ITS all right mama: Investigating the formation of chimeric sequences in the ITS2 region by DNA metabarcoding analyses of fungal mock communities of different complexities. *Molecular Ecology Resources*, *17*, 730–741. https://doi.org/10.1111/1755-0998.12622

Abarenkov, K., Somervuo, P., Nilsson, R. H., Kirk, P. M., Huotari, T., Abrego, N., & Ovaskainen, O. (2018). Protax-fungi: A web-based tool for probabilistic taxonomic placement of fungal internal transcribed spacer sequences. *New Phytologist*, *220*, 517–525.

Abarenkov, K., Tedersoo, L., Nilsson, R. H., Vellak, K., Saar, I., Veldre, V., Parmasto, E., Prous, M., Aan, A., Ots, M., Kurina, O., Ostonen, I., Jõgeva, J., Halapuu, S., Põldmaa, K., Toots, M., Truu, J., Larsson, K.-H., & Kõljalg, U. (2010). PlutoF – a web-based workbench for ecological and taxonomic research with an online implementation for fungal ITS sequences. *Evolutionary Bioinformatics*, *6*, 189–196. https://doi.org/10.4137/EBO.S6271

Abdala Asbun, A., Besseling, M. A., Balzano, S., vanBleijswijk, J. D., Witte, H. J., Villanueva, L., & Engelmann, J. C. (2020). Cascabel: A scalable and versatile amplicon sequence data analysis pipeline delivering reproducible and documented results. *Frontiers in Genetics*, *11*, 1329.

Ahn, J. H., Kim, B. Y., Song, J., & Weon, H. Y. (2012). Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *Journal of Microbiology*, *50*, 1071–1074. https://doi.org/10.1007/s12275-012-2642-z

Albanese, D., Fontana, P., De Filippo, C., Cavalieri, D., & Donati, C. (2015). MICCA: A complete and accurate software for taxonomic profiling of metagenomic data. *Scientific Reports*, *5*, 9743. https://doi.org/10.1038/srep09743

Alberdi, A., Aizpurua, O., Gilbert, M. T., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*, 134–147. https://doi.org/10.1111/2041-210X.12849

Alberdi, A., & Gilbert, M. T. (2019). A guide to the application of Hill numbers to DNA-based diversity analyses. *Molecular Ecology Resources*, *19*, 804–817. https://doi.org/10.1111/1755-0998.13014

Alteio, L. V., Séneca, J., Canarini, A., Angel, R., Jansa, J., Guseva, K., Kaiser, C., Richter, A., & Schmidt, H. (2021). A critical perspective on interpreting amplicon sequencing data in soil ecological research. *Soil Biology and Biochemistry*, *160*, 108357. https://doi.org/10.1016/j.soilbio.2021.108357

Amend, A., Burgaud, G., Cunliffe, M., Edgcomb, V. P., Ettinger, C. L., Gutiérrez, M. H., Heitman, J., Hom, E. F. Y., Ianiri, G., Jones, A. C., Kagami, M., Picard, K. T., Quandt, C. A., Raghukumar, S., Riquelme, M., Stajich, J., Vargas-Muñiz, J., Walker, A. K., Yarden, O., & Gladfelter, A. S. (2019). Fungi in the marine environment: Open questions and unsolved problems. *MBio*, *10*, e01189-18.https://doi.org/10.1128/mBio.01189-18

Amend, A. S., Seifert, K. A., & Bruns, T. D. (2010). Quantifying microbial communities with 454 pyrosequencing: Does read abundance count?*Molecular Ecology*, *19*, 5555–5565. https://doi.org/10.1111/j.1365-294X.2010.04898.x

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems*, *2*, e00191-16. https://doi.org/10.1128/mSystems.00191-16

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*, 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Anderson, M. J., Gorley, R. N., & Clarke, R. K. (2008). *Permanova+ for primer: Guide to software and statistical methods*. Primer-E.

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

Anslan, S., Li, H., Künzel, S., & Vences, M. (2021). Microbiomes from feces vs. gut in tadpoles: Distinct community compositions between substrates and preservation methods. *Salamandra*, *57*, 96–104. https://doi.org/10.1101/651612

Anslan, S., Nilsson, R. H., Wurzbacher, C., Baldrian, P., Tedersoo, L., & Bahram, M. (2018). Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. *MycoKeys*, *39*, 29–40. https://doi.org/10.3897/mycokeys.39.28109

Antich, A., Palacin, C., Wangensteen, O. S., & Turon, X. (2021). To denoise or to cluster, that is not the question: Optimizing pipelines for COI metabarcoding and metaphylogeography. *BMC Bioinformatics*, *22*, 177. https://doi.org/10.1186/s12859-021-04115-6

Antoninka, A., Wolf, J. E., Bowker, M., Classen, A. T., & Johnson, N. C. (2009). Linking above- and belowground responses to global change at community and ecosystem scales. *Global Change Biology*, *15*, 914–929. https://doi.org/10.1111/j.1365-2486.2008.01760.x

Bahram, M., Harend, H., & Tedersoo, L. (2014). Network perspectives of ectomycorrhizal associations. *Fungal Ecology*, *7*, 70–77. https://doi.org/10.1016/j.funeco.2013.10.003

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., Bengtsson-Palme, J., Anslan, S., Coelho, L. P., Harend, H., Huerta-Cepas, J., Medema, M. H., Maltz, M. R., Mundra, S., Olsson, P. A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., … Bork, P. (2018). Structure and function of the global topsoil microbiome. *Nature*, *560*, 233–237. https://doi.org/10.1038/s41586-018-0386-6

Bahram, M., Kõljalg, U., Courty, P.-E., Diédhiou, A. G., Kjøller, R., Põlme, S., Ryberg, M., Veldre, V., & Tedersoo, L. (2013). The distance-decay of similarity in communities of ectomycorrhizal fungi in different ecosystems and scales. *Journal of Ecology*, *101*, 1335–1344. https://doi.org/10.1111/1365-2745.12120

Bahram, M., Peay, K. G., & Tedersoo, L. (2015). Local-scale biogeography and spatiotemporal variability in communities of mycorrhizal fungi. *New Phytologist*, *205*, 1454–1463. https://doi.org/10.1111/nph.13206

Bakker, M. G. (2018). A fungal mock community control for amplicon sequencing experiments. *Molecular Ecology Resources*, *18*, 541–556. https://doi.org/10.1111/1755-0998.12760

Bálint, M., Bahram, M., Eren, A. M., Faust, K., Fuhrman, J. A., Lindahl, B., O'Hara, R. B., Öpik, M., Sogin, M. L., Unterseher, M., & Tedersoo, L. (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews*, *40*, 686–700. https://doi.org/10.1093/femsre/fuw017

Bálint, M., Pfenninger, M., Grossart, H.-P., Taberlet, P., Vellend, M., Leibold, M. A., Englund, G., & Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology & Evolution*, *33*, 945–957. https://doi.org/10.1016/j.tree.2018.09.003

Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., & Stamatakis, A. (2019). EPA-ng: Massively parallel evolutionary

placement of genetic sequences. *Systematic Biology*, *68*, 365–369. https://doi.org/10.1093/sysbio/syy054

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In: *International AAAI Conference on Weblogs and Social Media*. Association for the Advancement of Artificial Intelligence.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., Amend, A. S., Jumpponen, A., Unterseher, M., Kristiansson, E., Abarenkov, K., Bertrand, Y. J. K., Sanli, K., Eriksson, K. M., Vik, U., … Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution*, *4*, 914–919. https://doi.org/10.1111/2041-210X.12073

Berry, D., & Loy, A. (2018). Stable-isotope probing of human and animal microbiome function. *Trends in Microbiology*, *26*, 999–1007. https://doi.org/10.1016/j.tim.2018.06.004

Beule, L., & Karlovsky, P. (2020). Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): Application to microbial communities. *PeerJ*, *8*, e9593. https://doi.org/10.7717/peerj.9593

Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P. M., Reith, F., Dennis, P. G., Breed, M. F., Brown, B., Brown, M. V., Brugger, J., Byrne, M., Caddy-Retalic, S., Carmody, B., Coates, D. J., Correa, C., Ferrari, B. C., Gupta, V. V. S. R., Hamonts, K., Haslem, A., … Young, A. (2016). Introducing BASE: The Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience*, *5*, 21. https://doi.org/10.1186/s13742-016-0126-5

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, *23*, 1050–1063. https://doi.org/10.1111/ele.13525

Blazewicz, S. J., Barnard, R. L., Daly, R. A., & Firestone, M. K. (2013). Evaluating rRNA as an indicator of microbial activity in environmental communities: Limitations and uses. *The ISME Journal*, *7*, 2061–2068. https://doi.org/10.1038/ismej.2013.102

Bohmann, K., Elbrecht, V., Carøe, C., Bista, I., Leese, F., Bunce, M., Yu, D. W., Seymour, M., Dumbrell, A. J., & Creer, S. (2022). Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Molecular Ecology Resources*, *22*(4), 1231–1246https://doi.org/10.1111/1755-0998.13512

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*, 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bork, P., Bowler, C., De Vargas, C., Gorsky, G., Karsenti, E., & Wincker, P. (2015). Tara Oceans studies plankton at planetary scale. *Science*, *348*, 873. https://doi.org/10.1126/science.aac5605

Bowers, R. M., Clum, A., Tice, H., Lim, J., Singh, K., Ciobanu, D., Ngan, C. Y., Cheng, J.-F., Tringe, S. G., & Woyke, T. (2015). Impact of library preparation protocols and template quantity on the metagenomic reconstruction of a mock microbial community. *BMC Genomics*, *16*, 856. https://doi.org/10.1186/s12864-015-2063-6

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). Obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*, 176–182. https://doi.org/10.1111/1755-0998.12428

Brabcova, V., Novakova, M., Davidova, A., & Baldrian, P. (2016). Dead fungal mycelium in forest soil represents a decomposition hotspot and a habitat for a specific microbial community. *New Phytologist*, *210*, 1369–1381. https://doi.org/10.1111/nph.13849

Braid, M. D., Daniels, L. M., & Kitts, C. L. (2003). Removal of PCR inhibitors from soil DNA by chemical flocculation. *Journal of Microbiological Methods*, *52*, 389–393. https://doi.org/10.1016/S0167-7012(02)00210-5

Brandt, M. I., Trouche, B., Quintric, L., Günther, B., Wincker, P., Poulain, J., & Arnaud-Haond, S. (2021). Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. *Molecular Ecology Resources*, *21*, 1904–1921. https://doi.org/10.1111/1755-0998.13398

Bruns, T. D., & Taylor, J. W. (2016). Comment on "Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism". *Science*, *351*, 826. https://doi.org/10.1126/science.aad2346

Bunge, J., Willis, A., & Walsh, F. (2014). Estimating the number of species in microbial diversity studies. *Annual Review of Statistics and Its Application*, *1*, 427–445. https://doi.org/10.1146/annurev-statistics-022513-115654

Buschmann, T., & Bystrykh, L. V. (2013). Levenshtein error-correcting barcodes for multiplexed DNA sequencing. *BMC Bioinformatics*, *14*, 272. https://doi.org/10.1186/1471-2105-14-272

Buttigieg, P. L., & Ramette, A. (2014). A guide to statistical analysis in microbial ecology: A community-focused living review of multivariate data analyses. *FEMS Microbiology Ecology*, *90*, 543–550. https://doi.org/10.1111/1574-6941.12437

Byrne, A. Q., Rothstein, A. P., Poorten, T. J., Erens, J., Settles, M. L., & Rosenblum, E. B. (2017). Unlocking the story in the swab: A new genotyping assay for the amphibian chytrid fungus *Batrachochytrium dendrobatidis*. *Molecular Ecology Resources*, *17*, 1283–1292. https://doi.org/10.1111/1755-0998.12675

Caceres, M. D., & Legendre, P. (2009). Associations between species and groups of sites: Indices and statistical inference. *Ecology*, *90*, 3566–3574. https://doi.org/10.1890/08-1823.1

Cai, F., & Druzhinina, I. S. (2021). In honor of John Bissett: Authoritative guidelines on molecular identification of *Trichoderma*. *Fungal Diversity*, *107*, 1–69. https://doi.org/10.1007/s13225-020-00464-4

Calderon-Sanou, I., Münkemüller, T., Boyer, F., Zinger, L., & Thuiller, W. (2020). From environmental DNA sequences to ecological conclusions: How strong is the influence of methodological choices?*Journal of Biogeography*, *47*, 193–206. https://doi.org/10.1111/jbi.13681

Callahan, B. J., Grinevich, D., Thakur, S., Balamotis, M. A., & Yehezkel, T. B. (2021). Ultra-accurate microbial amplicon sequencing with synthetic long reads. *Microbiome*, *9*, 130. https://doi.org/10.1186/s40168-021-01072-3

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*, 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*, 581–584. https://doi.org/10.1038/nmeth.3869

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

Carlsen, T., Aas, A. B., Lindner, D., Vrålstad, T., Schumacher, T., & Kauserud, H. (2012). Don't make a mista(g)ke: Is tag switching an overlooked source of error in amplicon pyrosequencing studies?*Fungal Ecology*, *5*, 747–749. https://doi.org/10.1016/j.funeco.2012.06.003

Carøe, C., & Bohmann, K. (2020). Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *Molecular Ecology Resources*, *20*, 1620–1631. https://doi.org/10.1111/1755-0998.13227

Castaño, C., Parladé, J., Pera, J., De Aragón, J. M., Alday, J. G., & Bonet, J. A. (2016). Soil drying procedure affects the DNA quantification of *Lactarius vinosus* but does not change the fungal community composition. *Mycorrhiza*, *26*, 799–808. https://doi.org/10.1007/s00572-016-0714-3

Castle, S. C., Song, Z., Gohl, D. M., Gutknecht, J. L. M., Rosen, C. J., Sadowsky, M. J., Samac, D. A., & Kinkel, L. L. (2018). DNA template dilution impacts amplicon sequencing-based estimates of soil fungal diversity. *Phytobiomes*, *2*, 100–107. https://doi.org/10.1094/PBIOMES-09-17-0037-R

Chalmandrier, L., Pansu, J., Zinger, L., Boyer, F., Coissac, E., Génin, A., Gielly, L., Lavergne, S., Legay, N., Schilling, V., Taberlet, P., Münkemüller, T., & Thuiller, W. (2019). Environmental and biotic drivers of soil microbial β-diversity across spatial and phylogenetic scales. *Ecography*, *42*, 2144–2156. https://doi.org/10.1111/ecog.04492

Chambouvet, A., Monier, A., Maguire, F., Itoïz, S., del Campo, J., Elies, P., Edvardsen, B., Eikreim, W., & Richards, T. A. (2019). Intracellular infection of diverse diatoms by an evolutionary distinct relative of the Fungi. *Current Biology*, *29*, 4093–4101. https://doi.org/10.1016/j.cub.2019.09.074

Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, *84*, 45–67. https://doi.org/10.1890/13-0133.1

Chazdon, R. L., Chao, A., Colwell, R. K., Lin, S.-Y., Norden, N., Letcher, S. G., Clark, D. B., Finegan, B., & Arroyo, J. P. (2011). A novel statistical method for classifying habitat generalists and specialists. *Ecology*, *92*, 1332–1343. https://doi.org/10.1890/10-1345.1

Clasen, L. A., Detheridge, A. P., Scullion, J., & Griffith, G. W. (2020). Soil stabilisation for DNA metabarcoding of plants and fungi. Implications for sampling at remote locations or via third-parties. *Metabarcoding and Metagenomics*, *4*, e58365. https://doi.org/10.3897/mbmg.4.58365

Cline, L. C., Song, Z., Al-Ghalith, G. A., Knights, D., & Kennedy, P. G. (2017). Moving beyond de novo clustering in fungal community ecology. *New Phytologist*, *216*, 629–634. https://doi.org/10.1111/nph.14752

Collier, J. E. (2020). *Applied structural equation modeling using AMOS: Basic to advanced techniques*. Routledge Publishing.

Colwell, R. K., Mao, C. X., & Chang, J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology*, *85*, 2717–2727. https://doi.org/10.1890/03-0557

Compson, Z. G., Monk, W. A., Curry, C. J., Gravel, D., Bush, A., Baker, C. J., Manir, M. S. A., Riazanov, A., Hajibabaei, M., Shokralla, S., Gibson, J. F., Stefani, S., Wright, M. T. G., & Baird, D. J. (2018). Linking DNA metabarcoding and text mining to create network-based biomonitoring tools: A case study on boreal wetland macroinvertebrate communities. *Advances in Ecological Research*, *59*, 33–74.

Connor, N., Barberán, A., & Clauset, A. (2017). Using null models to infer microbial co-occurrence networks. *PLoS One*, *12*, e0176751. https://doi.org/10.1371/journal.pone.0176751

Cristescu, M. E., & Hebert, P. D. (2018). Uses and misuses of environmental DNA in biodiversity science and conservation. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 209–230. https://doi.org/10.1146/annurev-ecolsys-110617-062306

D'Amore, R., Ijaz, U. Z., Schirmer, M., Kenny, J. G., Gregory, R., Darby, A. C., Shakya, M., Podar, M., Quince, C., & Hall, N. (2016). A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. *BMC Genomics*, *17*, 55. https://doi.org/10.1186/s12864-015-2194-9

Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ, A., Burla, S., Diedhiou, A. G., Hiiesalu, I., Jairus, T., Johnson, N. C., Kane, A., Koorem, K., Kochar, M., Ndiaye, C., Pärtel, M., Reier, Ü., Saks, Ü., Singh, R., … Zobel, M. (2015). Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science*, *349*, 970–973. https://doi.org/10.1126/science.aab1161

deCarcer, D. A., Denman, S. E., McSweeney, C., & Morrison, M. (2011). Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing datasets from gut microbiomes. *Applied and Environmental Microbiology*, *77*, 8795–8798. https://doi.org/10.1128/aem.05491-11

Delavaux, C. S., Bever, J. D., Karppinen, E. M., & Bainard, L. D. (2020). Keeping it cool: Soil sample cold pack storage and DNA shipment up to 1 month does not impact metabarcoding results. *Ecology and Evolution*, *10*, 4652–4664. https://doi.org/10.1002/ece3.6219

Delavaux, C. S., Sturmer, S. L., Wagner, M. R., Schütte, U., Morton, J. B., & Bever, J. D. (2021). Utility of LSU for environmental sequencing of arbuscular mycorrhizal fungi: A new reference database and pipeline. *New Phytologist*, *229*, 3048–3052. https://doi.org/10.1111/nph.17080

Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D., Berdugo, M., Campbell, C. D., & Singh, B. K. (2016). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nature Communications*, *7*, 10541. https://doi.org/10.1038/ncomms10541

Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., & Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinformatics*, *13*, 113. https://doi.org/10.1186/1471-2105-13-113

Deshpande, V., Wang, Q., Greenfield, P., Charleston, M., Porras-Alfaro, A., Kuske, C. R., Cole, J. R., Midgley, D. J., & Tran-Dinh, N. (2016). Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia*, *108*, 1–5. https://doi.org/10.3852/14-293

Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., Holdaway, R. J., Lear, G., Makiola, A., Morales, S. E., & Powell, J. R. & Weaver, L. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular Ecology Resources*, *18*, 940–952. https://doi.org/10.1111/1755-0998.12907

Dormann, C. F., Fründ, J., Blüthgen, N., & Gruber, B. (2009). Indices, graphs and null models: analyzing bipartite ecological networks. *Open Ecology Journal*, *2*, 7–24. https://doi.org/10.2174/1874213000902010007

Douhan, G. W., Vincenot, L., Gryta, H., & Selosse, M.-A. (2011). Population genetics of ectomycorrhizal fungi: From current knowledge to emerging directions. *Fungal Biology*, *115*, 569–597. https://doi.org/10.1016/j.funbio.2011.03.005

Dray, S., Blanchet, G., Borcard, D., Guenard, G., Jombart, T., Larocque, G., Legendre, P., & Madi, N. & Wagner, H. H. (2018). *Package 'adespatial'*. Available:https://cran.microsoft.com/web/packages/adespatial/adespatial.pdf

Edgar, R. C. (2016). UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*, 2016, 081257. https://doi.org/10.1101/081257

Edgar, R. C. (2018). UNCROSS2: identification of cross-talk in 16S rRNA OTU tables. *BioRxiv*, 2018, 400762.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*, 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Egan, C. P., Rummel, A., Kokkoris, V., Klironomos, J., Lekberg, Y., & Hart, M. (2018). Using mock communities of arbuscular mycorrhizal fungi to evaluate fidelity associated with Illumina sequencing. *Fungal Ecology*, *33*, 52–64. https://doi.org/10.1016/j.funeco.2018.01.004

Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in low microbial biomass microbiome studies: Issues and recommendations. *Trends in Microbiology*, *27*, 105–117. https://doi.org/10.1016/j.tim.2018.11.003

Escudié, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., Maman, S., Hernandez-Raquet, G., Combes, S., & Pascal, G. (2018). FROGS: Find, rapidly, OTUs with galaxy solution. *Bioinformatics*, *34*, 1287–1294. https://doi.org/10.1093/bioinformatics/btx791

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, *43*, 2513–2524. https://doi.org/10.1093/nar/gkv107

Estensmo, E. L., Maurice, S., Morgado, L., Martin-Sanchez, P. M., Skrede, I., & Kauserud, H. (2021). The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species. *Molecular Ecology Resources*, *21*, 1141–1148. https://doi.org/10.1111/1755-0998.13329

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, *2*, 6. https://doi.org/10.1186/2049-2618-2-6

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: An updated review. *Ecological Processes*, *5*, 19. https://doi.org/10.1186/s13717-016-0063-3

Faust, K. (2021). Open challenges for microbial network construction and analysis. *The ISME Journal*, *15*, 3111–3118. https://doi.org/10.1038/s41396-021-01027-4

Faust, K., & Raes, J. (2016). CoNet app: Inference of biological association networks using Cytoscape. *F1000Research*, *5*, 1519. https://doi.org/10.12688/f1000research.9050.2

Fernandez, C. W., Nguyen, N. H., Stefanski, A., Han, Y., Hobbie, S. E., Montgomery, R. A., Reich, P. B., & Kennedy, P. G. (2017). Ectomycorrhizal fungal response to warming is linked to poor host performance at the boreal-temperate ecotone. *Global Change Biology*, *23*, 1598–1609. https://doi.org/10.1111/gcb.13510

Fischer, M., Renevey, N., Thür, B., Hoffmann, D., Beer, M., & Hoffmann, B. (2016). Efficacy assessment of nucleic acid decontamination reagents used in molecular diagnostic laboratories. *PLoS One*, *11*, e0159274. https://doi.org/10.1371/journal.pone.0159274

Forin, N., Nigris, S., Voyron, S., Girlanda, M., Vizzini, A., Casadoro, G., & Baldan, B. (2018). Next generation sequencing of ancient fungal specimens: the case of the Saccardo Mycological Herbarium. *Frontiers in Ecology and Evolution*, *6*, 129. https://doi.org/10.3389/fevo.2018.00129

Foster, Z. S. L., Sharpton, T. J., & Grünwald, N. J. (2017). Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Computational Biology*, *13*, e1005404. https://doi.org/10.1371/journal.pcbi.1005404

Fouquier, J., Rideout, J. R., Bolyen, E., Chase, J., Shiffer, A., McDonald, D., Knight, R., Caporaso, J. G., & Kelley, S. T. (2016). Ghost-tree: Creating hybrid-gene phylogenetic trees for diversity analyses. *Microbiome*, *4*, 11. https://doi.org/10.1186/s40168-016-0153-6

Friedman, J., & Alm, E. J. (2011). Inferring correlation networks from genomic survey data. *PLoS Computational Biology*, *8*, e1002687. https://doi.org/10.1371/journal.pcbi.1002687

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*, 1188. https://doi.org/10.1038/s41467-017-01312-x

Furneaux, B., Bahram, M., Rosling, A., Yorou, N. S., & Ryberg, M. (2021). Long- and short-read metabarcoding technologies reveal similar spatiotemporal structures in fungal communities. *Molecular Ecology Resources*, *21*, 1833–1849. https://doi.org/10.1111/1755-0998.13387

Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., Pantic, N., Admassu, T., James, P., Warland, A., Jordan, M., Ciccone, J., Serra, S., Keenan, J., Martin, S., McNeill, L., Wallace, E. J., Jayasinghe, L., Wright, C., … Turner, D. J. (2018). Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, *15*, 201–206. https://doi.org/10.1038/nmeth.4577

Geisen, S., Tveit, A. T., Clark, I. M., Richter, A., Svenning, M. M., Bonkowski, M., & Urich, T. (2015). Metatranscriptomic census of active protists in soils. *The ISME Journal*, *9*, 2178–2190. https://doi.org/10.1038/ismej.2015.30

Genuer, R., Poggi, J.-M., & Tuleau-Malot, C. (2019). *VSURF: variable selection using Random Forests. R package version 1.1.0*. https://CRAN.R-project.org/package=VSURF

Georgieva, D., Liu, Q., Wang, K., & Egli, D. (2020). Detection of base analogs incorporated during DNA replication by nanopore sequencing. *Nucleic Acids Research*, *48*, e88. https://doi.org/10.1093/nar/gkaa517

Gherbawy, Y., & Voigt, K. (Eds.) (2010). *Molecular identification of fungi*. Springer.

Ghori, N. U., Moreira-Grez, B., Vuong, P., Waite, I., Morald, T., Wise, M., & Whiteley, A. S. (2015). RNA stable isotope probing (RNA-SIP). *Methods in Molecular Biology*, *2046*, 31–44.

Glassman, S. I., & Martiny, J. B. (2018). Ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *mSphere*, *3*, e00148-18. https://doi.org/10.1128/mSphere.00148-18

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome datasets are compositional: And this is not optional. *Frontiers in Microbiology*, *8*, 2224. https://doi.org/10.3389/fmicb.2017.02224

Gohl, D. M., Auch, B., Certano, A., LeFrançois, B., Bouevitch, A., Doukhanine, E., Fragel, C., Macklaim, J., Hollister, E., Garbe, J., & Beckman, K. B. (2021). Dissecting and tuning primer editing by proofreading polymerases. *Nucleic Acids Research*, *49*, e87. https://doi.org/10.1093/nar/gkab471

Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T. J., Clayton, J. B., Johnson, T. J., Hunter, R., Knights, D., & Beckman, K. B. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nature Biotechnology*, *34*, 942–949. https://doi.org/10.1038/nbt.3601

Gotelli, N. J., & Ellison, A. M. (2013). *A primer of ecological statistics*. Sinauer Associates.

Grossart, H. P., Van den Wyngaert, S., Kagami, M., Wurzbacher, C., Cunliffe, M., & Rojas-Jimenez, K. (2019). Fungi in aquatic ecosystems. *Nature Reviews Microbiology*, *17*, 339–354. https://doi.org/10.1038/s41579-019-0175-8

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*, *30*, 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Guerrieri, A., Bonin, A., Münkemüller, T., Gielly, L., Thuiller, W., & Ficetola, G. F. (2021). Effects of soil preservation for biodiversity monitoring using environmental DNA. *Molecular Ecology*, *30*, 3313–3325. https://doi.org/10.1111/mec.15674

Gweon, H. S., Oliver, A., Taylor, J., Booth, T., Gibbs, M., Read, D. S., Griffiths, R. I. & Schonrogge, K. (2015). PIPITS: An automated pipeline for analyses of fungal internal transcribed spacer sequences from the Illumina sequencing platform. *Methods in Ecology and Evolution*, *6*, 973–980. https://doi.org/10.1111/2041-210x.12399

Halwachs, B., Madhusudhan, N., Krause, R., Nilsson, R. H., Moissl-Eichinger, C., Högenauer, C., Thallinger, G. G., & Gorkiewicz, G. (2017). Critical issues in mycobiota analysis. *Frontiers in Microbiology*, *8*, 180. https://doi.org/10.3389/fmicb.2017.00180

Hannula, S. E., Morriën, E., deHollander, M., van derPutten, W. H., vanVeen, J. A., & deBoer, W. (2017). Shifts in rhizosphere fungal community during secondary succession following abandonment from agriculture. *The ISME Journal*, *11*, 2294–2304. https://doi.org/10.1038/ismej.2017.90

Hanson, C. A., Allison, S. D., Bradford, M. A., Wallenstein, M. D., & Treseder, K. K. (2008). Fungal taxa target different carbon sources in forest soil. *Ecosystems*, *11*, 1157–1167. https://doi.org/10.1007/s10021-008-9186-4

Harris, J. K., Sahl, J. W., Castoe, T. A., Wagner, B. D., Pollock, D. D., & Spear, J. R. (2010). Comparison of normalization methods for construction of large, multiplex amplicon pools for next-generation sequencing. *Applied and Environmental Microbiology*, *76*, 3863–3868. https://doi.org/10.1128/AEM.02585-09

He, Y., Caporaso, J. G., Jiang, X.-T., Sheng, H.-F., Huse, S. M., Rideout, J. R., Edgar, R. C., Kopylova, E., Walters, W. A., Knight, R., & Zhou, H.-W.

(2015). Stability of operational taxonomic units: An important but neglected property for analyzing microbial diversity. *Microbiome*, *3*, 20. https://doi.org/10.1186/s40168-015-0081-x

Heeger, F., Bourne, E. C., Baschien, C., Yurkov, A., Bunk, B., Spröer, C., Overmann, J., & Mazzoni, C. J. & Monaghan, M. T. (2018). Long-read DNA metabarcoding of ribosomal rRNA in the analysis of fungi from aquatic environments. *Molecular Ecology Resources*, *18*, 1500–1514. https://doi.org/10.1111/1755-0998.12937

Holm, J. B., Humphrys, M. S., Robinson, C. K., Settles, M. L., Ott, S., Fu, L., & Ravel, J. (2020). Ultrahigh-throughput multiplexing and sequencing of >500-base-pair amplicon regions on the illumina HiSeq 2500 platform. *mSystems*, *4*, e00029-19. https://doi.org/10.1128/mSystems.00029-19

Horn, S., Caruso, T., Verbruggen, E., Rillig, M. C., & Hempel, S. (2014). Arbuscular mycorrhizal fungal communities are phylogenetically clustered at small scales. *The ISME Journal*, *8*, 2231–2242. https://doi.org/10.1038/ismej.2014.72

Hugerth, L. W., & Andersson, A. F. (2017). Analysing microbial community composition through amplicon sequencing: From sampling to hypothesis testing. *Frontiers in Microbiology*, *8*, 1561. https://doi.org/10.3389/fmicb.2017.01561

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K. E., & Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology*, *83*, 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Jalili, V., Afgan, E., Gu, Q., Clements, D., Blankenberg, D., Goecks, J., Taylor, J., & Nekrutenko, A. (2020). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Research*, *48*, W395–W402. https://doi.org/10.1093/nar/gkaa434

Jarman, S. N., Berry, O., & Bunce, M. (2018). The value of environmental DNA biobanking for long-term biomonitoring. *Nature Ecology & Evolution*, *2*, 1192–1193. https://doi.org/10.1038/s41559-018-0614-3

Joos, L., Beirinckx, S., Haegeman, A., Debode, J., Vandecasteele, B., Baeyen, S., Goormachtig, S., Clement, L., & De Tender, C. (2020). Daring to be differential: Metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. *BMC Genomics*, *21*, 733. https://doi.org/10.1186/s12864-020-07126-4

Jumpponen, A., & Jones, K. L. (2009). Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. *New Phytologist*, *184*, 438–448. https://doi.org/10.1111/j.1469-8137.2009.02990.x

Karst, S. M., Ziels, R. M., Kirkegaard, R. H., Sørensen, E. A., McDonald, D., Zhu, Q., Knight, R., & Albertsen, M. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with nanopore or PacBio sequencing. *Nature Methods*, *18*, 165–169. https://doi.org/10.1038/s41592-020-01041-y

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*, 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Kennedy, P. G., Cline, L. C., & Song, Z. (2018). Probing promise versus performance in longer read fungal metabarcoding. *New Phytologist*, *217*, 973–976. https://doi.org/10.1111/nph.14883

Kolaříková, Z., Slavíková, R., Krüger, C., Krüger, M., & Kohout, P. (2021). PacBio sequencing of *Glomeromycota* rDNA: A novel amplicon covering all widely used ribosomal barcoding regions and its applicability in taxonomy and ecology of arbuscular mycorrhizal fungi. *New Phytologist*, *231*, 490–499. https://doi.org/10.1111/nph.17372

Kõljalg, U., Nilsson, H. R., Schigel, D., Tedersoo, L., Larsson, K.-H., May, T. W., Taylor, A. F. S., Jeppesen, T. S., Frøslev, T. G., Lindahl, B. D., Põldmaa, K., Saar, I., Suija, A., Savchenko, A., Yatsiuk, I., Adojaan, K., Ivanov, F., Piirmann, T., Pöhönen, R., … Abarenkov, K. (2020). The Taxon Hypothesis paradigm—on the unambiguous detection and communication of taxa. *Microorganisms*, *8*, 1910. https://doi.org/10.3390/microorganisms8121910

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

Kõljalg, U., Tedersoo, L., Nilsson, R. H., & Abarenkov, K. (2016). Digital identifiers for fungal species. *Science*, *352*, 1182–1183. https://doi.org/10.1126/science.aaf7115

Kumar, G., Eble, J. E., & Gaither, M. R. (2020). A practical guide to sample preservation and pre-PCR processing of aquatic environmental DNA. *Molecular Ecology Resources*, *20*, 29–39. https://doi.org/10.1111/1755-0998.13107

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., & Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLoS Computational Biology*, *11*, e1004226. https://doi.org/10.1371/journal.pcbi.1004226

Kyaschenko, Y., Clemmensen, K., Hagenbo, A., Karltun, E., & Lindahl, B. (2017). Shift in fungal communities and associated enzyme activities along an age gradient of managed *Pinus sylvestris* stands. *The ISME Journal*, *11*, 863–874. https://doi.org/10.1038/ismej.2016.184

Lagerborg, K. A., Normandin, E., Bauer, M. R., Adams, G., Figueroa, K., Loreth, C., Gladden-Young, A., Shaw, B., Pearlman, L., Shenoy, E. S., Hooper, D., Pierce, V. M., Zachary, K. C., Park, D. J., MacInnis, B. L., Lemieux, J. E., Sabeti, P. C., Reilly, S. K., & Siddle, K. J. (2021). DNA spike-ins enable confident interpretation of SARS-CoV-2 genomic data from amplicon-based sequencing. *bioRxiv*, 2021, 435654. doi: https://doi.org/10.1101/2021.03.16.435654

Lai, J., Zou, Y., Zhang, J., & Peres-Neto, P. R. (2022). Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca. hp R package. *Methods in Ecology and Evolution*, in press. doi: https://doi.org/10.1111/2041-210X.13800

Larsson, A. J., Stanley, G., Sinha, R., Weissman, I. L., & Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. *Nature Methods*, *15*, 305–307. https://doi.org/10.1038/nmeth.4666

Lefcheck, J. S. (2016). piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, *7*, 573–579.

Legendre, P., & Legendre, L. (2012). *Numerical ecology*. Elsevier.

Leite, M. F., & Kuramae, E. E. (2020). You must choose, but choose wisely: Model-based approaches for microbial community analysis. *Soil Biology and Biochemistry*, *151*, 108042.

Lennon, J. T., Muscarella, M. E., Placella, S. A., & Lehmkuhl, B. K. (2018). How, when, and where relic DNA affects microbial diversity. *MBio*, *9*, e00637–e718. https://doi.org/10.1128/mBio.00637-18

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*, W293–W296. https://doi.org/10.1093/nar/gkab301

Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M., Barker, K. B., Baumgartner, B., Belov, K., Bertorelle, G., Blaxter, M. L., Cai, J., Caperello, N. D., Carlson, K., Castilla-Rubio, J. C., Chaw, S.-M., Chen, L., Childers, A. K., Coddington, J. A., … Zhang, G. (2022). The Earth BioGenome Project 2020: Starting the Clock. *Proceedings of the National Academy of Sciences, USA*, *119*, e2115635118. https://doi.org/10.1073/pnas.2115635118

Liaw, A., & Wiener, A. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Lin, H., & Peddada, S. D. (2020a). Analysis of compositions of microbiomes with bias correction. *Nature Communications*, *11*, 3514. https://doi.org/10.1038/s41467-020-17041-7

Lin, H., & Peddada, S. D. (2020b). Analysis of microbial compositions: A review of normalization and differential abundance analysis. *NPJ Biofilms and Microbiomes*, *6*, 60. https://doi.org/10.1038/s41522-020-00160-w

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., Kõljalg, U., Pennanen, T., Rosendahl, S., Stenlid, J., & Kauserud, H. (2013). Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New Phytologist*, *199*, 288–299. https://doi.org/10.1111/nph.12243

Lindner, D. L., Carlsen, T., Nilsson, R. H., Davey, M., Schumacher, T., & Kauserud, H. (2013). Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution*, *3*, 1751–1764. https://doi.org/10.1002/ece3.586

Liu, C., Cui, Y., Li, X., & Yao, M. (2021). Microeco: An R package for data mining in microbial community ecology. *FEMS Microbiology Ecology*, *97*, fiaa255. https://doi.org/10.1093/femsec/fiaa255

Lofgren, L. A., Uehling, J. K., Branco, S., Bruns, T. D., Martin, F., & Kennedy, P. G. (2019). Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Molecular Ecology*, *28*, 721–730. https://doi.org/10.1111/mec.14995

Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, R., Drenkhan, R., & Tedersoo, L. (2019). Relative Performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) third-generation sequencing instruments in identification of agricultural and forest fungal pathogens. *Applied and Environmental Microbiology*, *85*, e01368–e1419. https://doi.org/10.1128/AEM.01368-19

Loos, D., Zhang, L., Beemelmanns, C., Kurzai, O., & Panagiotou, G. (2021). DAnIEL: A user-friendly web server for fungal ITS amplicon sequencing data. *Frontiers in Microbiology*, *12*, https://doi.org/10.3389/fmicb.2021.720513. 720513.

Lopez-Mondejar, R., Tláskal, V., Větrovský, T., Štursová, M., Toscan, R., daRocha, U. N., & Baldrian, P. (2020). Metagenomics and stable isotope probing reveal the complementary contribution of fungal and bacterial communities in the recycling of dead biomass in forest soil. *Soil Biology and Biochemistry*, *148*, 107875https://doi.org/10.1016/j.soilbio.2020.107875

Lücking, R., Aime, M. C., Robbertse, B., Miller, A. N., Aoki, T., Ariyawansa, H. A., Cardinali, G., Crous, P. W., Druzhinina, I. S., Geiser, D. M., Hawksworth, D. L., Hyde, K. D., Irinyi, L., Jeewon, R., Johnston, P. R., Kirk, P. M., Malosso, E., May, T. W., Meyer, W., … Schoch, C. L. (2021). Fungal taxonomy and sequence-based nomenclature. *Nature Microbiology*, *6*, 540–548. https://doi.org/10.1038/s41564-021-00888-x

Maeda, T., Kobayashi, Y., Kameoka, H., Okuma, N., Takeda, N., Yamaguchi, K., Bino, T., Shigenobu, S., & Kawaguchi, M. (2018). Evidence of non-tandemly repeated rDNAs and their intragenomic heterogeneity in *Rhizophagus irregularis*. *Communications Biology*, *1*, 87. https://doi.org/10.1038/s42003-018-0094-7

Mahe, F., Czech, L., Stamatakis, A., Quince, C., deVargas, C., Dunthorn, M., & Rognes, T. (2022). Swarm v3: Towards tera-scale amplicon clustering. *Bioinformatics*, *38*, 267–269. https://doi.org/10.1093/bioinformatics/btab493

Manion, G., Lisk, M., Ferrier, S., Nieto-Lugilde, D., Mokany, K., & Fitzpatrick, M. C. (2018). *Gdm: Generalized Dissimilarity Modeling*. Available: https://CRAN.R-project.org/package=gdm

Martino, C., Morton, J. T., Marotz, C. A., Thompson, L. R., Tripathi, A., Knight, R., & Zengler, K. (2019). A novel sparse compositional technique reveals microbial perturbations. *Msystems*, *4*, e00016–19. https://doi.org/10.1128/mSystems.00016-19

Matchado, M. S., Lauber, M., Reitmeier, S., Kacprowski, T., Baumbach, J., Haller, D., & List, M. (2021). Network analysis methods for studying microbial communities: A mini review. *Computational and Structural Biotechnology Journal*, *19*, 2687–2698. https://doi.org/10.1016/j.csbj.2021.05.001

Matsuoka, S., Sugiyama, Y., Sato, H., Katano, I., Harada, K., & Doi, H. (2019). Spatial structure of fungal DNA assemblages revealed with eDNA metabarcoding in a forest river network in western Japan. *Metabarcoding and Metagenomics*, *3*, e36335. https://doi.org/10.3897/mbmg.3.36335

McKnight, D. T., Huerlimann, R., Bower, D. S., Schwarzkopf, L., Alford, R. A., & Zenger, K. R. (2019). MicroDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environmental DNA*, *1*, 14–25. https://doi.org/10.1002/edn3.11

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, *8*, e61217. https://doi.org/10.1371/journal.pone.0061217

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

Metsalu, T., & Vilo, J. (2015). ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research*, *43*, W566–W570. https://doi.org/10.1093/nar/gkv468

Mikryukov, V. S., Dulya, O. V., Likhodeevskii, G. A., & Vorobeichik, E. L. (2021). Analysis of ecological networks in multicomponent communities of microorganisms: Possibilities, limitations, and potential errors. *Russian Journal of Ecology*, *52*, 188–200. https://doi.org/10.1134/S1067413621030085

Nagler, M., Insam, H., Pietramellara, G., & Ascher-Jenull, J. (2018). Extracellular DNA in natural environments: Features, relevance and applications. *Applied Microbiology and Biotechnology*, *102*, 6343–6356. https://doi.org/10.1007/s00253-018-9120-4

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S., & Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community data sets by ecological guild. *Fungal Ecology*, *20*, 241–248. https://doi.org/10.1016/j.funeco.2015.06.006

Niku, J., Hui, F. K., Taskinen, S., & Warton, D. I. (2019). gllvm: Fast analysis of multivariate abundance data with generalized linear latent variable models in R. *Methods in Ecology and Evolution*, *10*, 2173–2182.

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., & Tedersoo, L. (2018). Mycobiome diversity: High-throughput sequencing and identification of fungi. *Nature Reviews in Microbiology*, *17*, 95–109. https://doi.org/10.1038/s41579-018-0116-y

Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., Tedersoo, L., Saar, I., Kõljalg, U., & Abarenkov, K. (2019). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research*, *47*, D259–D264. https://doi.org/10.1093/nar/gky1022

Nilsson, R. H., Taylor, A. F., Adams, R. I., Baschien, C., Bengtsson-Palme, J., Cangren, P., & Abarenkov, K. (2017). Taxonomic annotation of public fungal ITS sequences from the built environment–a report from an April 10–11, 2017 workshop (Aberdeen, UK). *MycoKeys*, *28*, 65.

Nilsson, R. H., Tedersoo, L., Abarenkov, K., Ryberg, M., Kristiansson, E., Hartmann, M., Schoch, C. L., Nylander, J. A. A., Bergsten, J., Porter, T. M., Jumpponen, A., Vaishampayan, P., Ovaskainen, O., Hallenberg, N., Bengtsson-Palme, J., Eriksson, K. M., Larsson, K.-H., Larsson, E., & Kõljalg, U. (2012). Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys*, *4*, 37–63. https://doi.org/10.3897/mycokeys.4.3606

Noble, W. S. (2009). How does multiple testing correction work?*Nature Biotechnology*, *27*, 1135–1137. https://doi.org/10.1038/nbt1209-1135

O'Donnell, K., Ward, T. J., Robert, V. A., Crous, P. W., Geiser, D. M., & Kang, S. (2015). DNA sequence-based identification of *Fusarium*: Current status and future directions. *Phytoparasitica*, *43*, 583–595. https://doi.org/10.1007/s12600-015-0484-z

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Peter Solymos, M., Stevens, H. H., Szoecs, E., & Wagner, H. (2019). *Vegan: Community Ecology Package*. Available:https://CRAN.R-project.org/package=vegan

Oliver, A. K., Brown, S. P., Callaham, M. A., & Jumpponen, A. (2015). Polymerase matters: Non-proofreading enzymes inflate fungal community richness estimates by up to 15%. *Fungal Ecology*, *15*, 86–89. https://doi.org/10.1016/j.funeco.2015.03.003

Oliveros, J. (2007). *An Interactive Tool for Comparing Lists with Venn Diagrams*. Available: http://bioinfogp.cnb.csic.es/tools/venny/index.html

Ondov, B. D., Bergman, N. H., & Philippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, *12*, 385. https://doi.org/10.1186/1471-2105-12-385

Öpik, M., Vanatoa, A., Vanatoa, E., Moora, M., Davison, J., Kalwij, J. M., Reier, Ü., & Zobel, M. (2010). The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (*Glomeromycota*). *New Phytologist*, *188*, 223–241. https://doi.org/10.1111/j.1469-8137.2010.03334.x

Ozkurt, E., Fritscher, J., Soranzo, N., & Ng, D. Y. (2021). LotuS2: An ultrafast and highly accurate tool for amplicon sequencing analysis. *bioRxiv*, *2021*, 474111. https://doi.org/10.1101/2021.12.24.474111

Paliy, O., & Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology*, *25*, 1032–1057. https://doi.org/10.1111/mec.13536

Palmer, J. M., Jusino, M. A., Banik, M. T., & Lindner, D. L. (2018). Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ*, *6*, e4925. https://doi.org/10.7717/peerj.4925

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, e6876. https://doi.org/10.7717/peerj.6876

Perez-Izquierdo, L., Zabal-Aguirre, M., González-Martínez, S. C., Buée, M., Verdú, M., Rincón, A., & Goberna, M. (2019). Plant intraspecific variation modulates nutrient cycling through its below ground rhizospheric microbiome. *Journal of Ecology*, *107*, 1594–1605. https://doi.org/10.1111/1365-2745.13202

Perez-Lamarque, B., Öpik, M., Maliet, O., Silva, A. C., Selosse, M. A., Martos, F., & Morlon, H. (2020). Global drivers of obligate mycorrhizal symbionts diversification. *bioRxiv*, *2020*, 224790.

Pinheiro, J., Bates, D., & DebRoy, S. (2011). *Nlme: Linear and nonlinear mixed effects models*. Available: http://cran.rproject.org/web/packages/nlme/

Pinol, J., Senar, M. A., & Symondson, W. O. (2019). The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. *Molecular Ecology*, *28*, 407–419. https://doi.org/10.1111/mec.14776

Piombo, E., Abdelfattah, A., Droby, S., Wisniewski, M., Spadaro, D., & Schena, L. (2021). Metagenomics approaches for the detection and surveillance of emerging and recurrent plant pathogens. *Microorganisms*, *9*, 188. https://doi.org/10.3390/microorganisms9010188

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., & Haustein, S. (2018). The state of OA: A large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, *6*, e4375. https://doi.org/10.7717/peerj.4375

Poelen, J. H., Simons, J. D., & Mungall, C. J. (2014). Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics*, *24*, 148–159. https://doi.org/10.1016/j.ecoinf.2014.08.005

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

Porras-Alfaro, A., Liu, K. L., Kuske, C. R., & Xie, G. (2014). From genus to phylum: Large-subunit and internal transcribed spacer rRNA operon regions show similar classification accuracies influenced by database composition. *Applied and Environmental Microbiology*, *80*, 829–840. https://doi.org/10.1128/aem.02894-13

Porter, T. M., & Hajibabaei, M. (2020). Putting COI metabarcoding in context: the utility of exact sequence variants (ESVs) in biodiversity analysis. *Frontiers in Ecology and Evolution*, *8*, 248. https://doi.org/10.3389/fevo.2020.00248

Porter, T. M., & Hajibabaei, M. (2021). Profile hidden Markov model sequence analysis can help remove putative pseudogenes from DNA barcoding and metabarcoding datasets. *BMC Bioinformatics*, *22*, 256. https://doi.org/10.1186/s12859-021-04180-x

Qian, H., & Jin, Y. (2016). An updated megaphylogeny of plants, a tool for generating plant phylogenies and an analysis of phylogenetic community structure. *Journal of Plant Ecology*, *9*, 233–239. https://doi.org/10.1093/jpe/rtv047

Qu, K., Guo, F., Liu, X., Lin, Y., & Zou, Q. (2019). Application of machine learning in microbiology. *Frontiers in Microbiology*, *10*, 827. https://doi.org/10.3389/fmicb.2019.00827

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*, D590–D596. https://doi.org/10.1093/nar/gks1219

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, *35*, 833–844. https://doi.org/10.1038/nbt.3935

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., & Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *GigaScience*, *8*, giz107. https://doi.org/10.1093/gigascience/giz107

Rajala, T., Peltoniemi, M., Hantula, J., Mäkipää, R., & Pennanen, T. (2011). RNA reveals a succession of active fungi during the decay of Norway spruce logs. *Fungal Ecology*, *4*, 437–448. https://doi.org/10.1016/j.funeco.2011.05.005

Rammitsu, K., Kajita, T., Imai, R., & Ogura-Tsujita, Y. (2021). Strong primer bias for Tulasnellaceae fungi in metabarcoding: Specific primers improve the characterization of the mycorrhizal communities of epiphytic orchids. *Mycoscience*, *62*, MYC551. https://doi.org/10.47371/mycosci.2021.06.005

Rao, C., Coyte, K. Z., Bainter, W., Geha, R. S., Martin, C. R., & Rakoff-Nahoum, S. (2021). Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature*, *591*, 633–638. https://doi.org/10.1038/s41586-021-03241-8

Rillig, M. C., Ryo, M., Lehmann, A., Aguilar-Trigueros, C. A., Buchert, S., Wulf, A., Iwasaki, A., Roy, J., & Yang, G. (2019). The role of multiple global change factors in driving soil functions and microbial biodiversity. *Science*, *366*, 886–890. https://doi.org/10.1126/science.aay2832

Rivers, A. R., Weber, K. C., Gardner, T. G., Liu, S., & Armstrong, S. D. (2018). ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis. *F1000Research*, *7*, 1418. https://doi.org/10.12688/f1000research.15704.1

Röttjers, L., & Faust, K. (2018). From hairballs to hypotheses – biological insights from microbial networks. *FEMS Microbiology Reviews*, *42*, 761–780. https://doi.org/10.1093/femsre/fuy030

Roy, J., Mazel, F., Sosa-Hernández, M. A., Dueñas, J. F., Hempel, S., Zinger, L., & Rillig, M. C. (2019). The relative importance of ecological drivers of arbuscular mycorrhizal fungal distribution varies with taxon phylogenetic resolution. *New Phytologist*, *224*, 936–948. https://doi.org/10.1111/nph.16080

Runnel, K., Abarenkov, K., Copot, O., Mikryukov, V., Kõljalg, U., Saar, I., & Tedersoo, L. (2022). DNA barcoding of fungal specimens using long-read high-throughput sequencing. *bioRxiv*, *2022*, 479507. doi: https://doi.org/10.1101/2022.02.08.479507

Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, *17*, e00547. https://doi.org/10.1016/j.gecco.2019.e00547

Sato, H., Sogo, Y., Doi, H., & Yamanaka, H. (2017). Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. *Scientific Reports*, *7*, 14860. https://doi.org/10.1038/s41598-017-14978-6

Sato, M. P., Ogura, Y., Nakamura, K., Nishida, R., Gotoh, Y., Hayashi, M., Hisatsune, J., Sugai, M., Takehiko, I., & Hayashi, T. (2019). Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. *DNA Research*, *26*, 391–398. https://doi.org/10.1093/dnares/dsz017

Schloss, P. D. (2020). Reintroducing mothur: 10 years later. *Applied and Environmental Microbiology*, *86*, e02343–e2419. https://doi.org/10.1128/AEM.02343-19

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*, 1289–1303. https://doi.org/10.1111/1755-0998.12402

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

Schwarzenbach, K., Enkerli, J., & Widmer, F. (2007). Objective criteria to assess representativity of soil fungal community profiles. *Journal of Microbiological Methods*, *68*, 358–366. https://doi.org/10.1016/j.mimet.2006.09.015

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*, 2498–2504. https://doi.org/10.1101/gr.1239303

Simmons, D. R., Bonds, A. E., Castillo, B. T., Clemons, R. A., Glasco, A. D., Myers, J. M., Thapa, N., Letcher, P. M., Powell, M. J., Longcore, J. E., & James, T. Y. (2020). The Collection of Zoosporic Eufungi at the University of Michigan (CZEUM): introducing a new repository of barcoded *Chytridiomyceta* and *Blastocladiomycota* cultures. *IMA Fungus*, *11*, 1–22. https://doi.org/10.1186/s43008-020-00041-z

Singer, E., Wagner, M., & Woyke, T. (2017). Capturing the genetic makeup of the active microbiome in situ. *The ISME Journal*, *11*, 1949–1963. https://doi.org/10.1038/ismej.2017.59

Song, Z., Schlatter, D., Kennedy, P., Kinkel, L. L., Kistler, H. C., Nguyen, N., & Bates, S. T. (2015). Effort versus reward: Preparing samples for fungal community characterization in high-throughput sequencing surveys of soils. *PLoS One*, *10*, e0127234. https://doi.org/10.1371/journal.pone.0127234

Stajich, J. E., Berbee, M. L., Blackwell, M., Hibbett, D. S., James, T. Y., Spatafora, J. W., & Taylor, J. W. (2009). The Fungi. *Current Biology*, *19*, R840–R845. https://doi.org/10.1016/j.cub.2009.07.004

Straub, D., Blackwell, N., Langarica-Fuentes, A., Peltzer, A., Nahnsen, S., & Kleindienst, S. (2020). Interpretations of environmental microbial community studies are biased by the selected 16S rRNA (gene) amplicon sequencing pipeline. *Frontiers in Microbiology*, *11*, 2652. https://doi.org/10.3389/fmicb.2020.550420

Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on! A new, conditional variable-importance measure for random forests available in the party package. *R Journal*, *1*, 14–17. https://doi.org/10.32614/RJ-2009-013

Sun, X., Hu, Y. H., Wang, J., Fang, C., Li, J., Han, M., & Gong, M. (2021). Efficient and stable metabarcoding sequencing data using a DNBSEQ-G400 sequencer validated by comprehensive community analyses. *Gigabyte*, *2021*, 1–15. https://doi.org/10.1101/2020.07.02.185710

Sze, M. A., & Schloss, P. D. (2019). The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. *mSphere*, *26*, e00163-19. https://doi.org/10.1128/mSphere.00163-19

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA: For Biodiversity Research and Monitoring*. Oxford University Press.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*, 2045–2050. https://doi.org/10.1111/j.1365-294x.2012.05470.x

Talas, L., Stivrins, N., Veski, S., Tedersoo, L., & Kisand, V. (2021). Sedimentary ancient DNA (sedaDNA) reveals fungal diversity and environmental drivers of community changes throughout the holocene in the present boreal Lake Lielais Svētiņu (Eastern Latvia). *Microorganisms*, *9*, 719. https://doi.org/10.3390/microorganisms9040719

Tanabe, A. S., & Toju, H. (2013). Two new computational methods for universal DNA Barcoding: A benchmark using barcode sequences of bacteria, archaea, animals, fungi, and land plants. *PLoS One*, *8*, e76910. https://doi.org/10.1371/journal.pone.0076910

Taylor, D. L., Hollingsworth, T. N., McFarland, J. W., Lennon, N. J., Nussbaum, C., & Ruess, R. W. (2014). A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological Monographs*, *84*, 3–20. https://doi.org/10.1890/12-1693.1

Tedersoo, L., Albertsen, M., Anslan, S., & Callahan, B. (2021). Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Applied and Environmental Microbiology*, *87*, e00626–e721. https://doi.org/10.1128/aem.00626-21

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

Tedersoo, L., Anslan, S., Bahram, M., Kõljalg, U., & Abarenkov, K. (2020). Identifying the 'unidentified'fungi: A global-scale long-read third-generation sequencing approach. *Fungal Diversity*, *103*, 273–293. https://doi.org/10.1007/s13225-020-00456-4

Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., Kõljalg, U., Kisand, V., Nilsson, H., Hildebrand, F., Bork, P., & Abarenkov, K. (2015). Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys*, *10*, 1–43. https://doi.org/10.3897/mycokeys.10.4852

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

Tedersoo, L., Drenkhan, R., Anslan, S., Morales-Rodrigues, C., & Cleary, M. (2019). High-throughput identification and diagnostics of pathogens and pests: overview and practical recommendations. *Molecular Ecology Resources*, *19*, 47–76. https://doi.org/10.1111/1755-0998.12959

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

Tedersoo, L., Mett, M., Ishida, T. A., & Bahram, M. (2013). Phylogenetic relationships among host plants explain differences in fungal

species richness and community composition in ectomycorrhizal symbiosis. *New Phytologist*, *199*, 822–831. https://doi.org/10.1111/nph.12328

Tedersoo, L., Mikryukov, V., Anslan, S., Bahram, M., Khalid, A. N., Corrales, A., Agan, A., Vasco-Palacios, A.-M., Saitta, A., Antonelli, A., Rinaldi, A. C., Verbeken, A., Sulistyo, B. P., Tamgnoue, B., Furneaux, B., Ritter, C. D., Nyamukondiwa, C., Sharp, C., Marín, C., … Abarenkov, K. (2021). The Global Soil Mycobiome consortium dataset for boosting fungal diversity research. *Fungal Diversity*, *111*, 573–588. https://doi.org/10.1007/s13225-021-00493-7

Tedersoo, L., Mikryukov, V., Zizka, A., Bahram, M., Hagh-Doust, N., Anslan, S., Bahram, M., Drenkhan, R., Pritsch, K., Buegger, F., Padari, A., Hagh-Doust, N., Mikryukov, V., Kõljalg, U., & Abarenkov, K. (2022). Endemicity and global change vulnerability of fungi. *bioRxiv*, *2022*, 484796. doi: https://doi.org/10.1101/2022.03.17.484796

Tedersoo, L., Sánchez-Ramírez, S., Kõljalg, U., Bahram, M., Döring, M., Schigel, D., May, T., Ryberg, M., & Abarenkov, K. (2018). High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Diversity*, *90*, 135–159. https://doi.org/10.1007/s13225-018-0401-0

Tedersoo, L., Tooming-Klunderud, A., & Anslan, S. (2018). PacBio metabarcoding of fungi and other eukaryotes: Biases and perspectives. *New Phytologist*, *217*, 1370–1385.

Tipton, L., Müller, C. L., Kurtz, Z. D., Huang, L., Kleerup, E., Morris, A., Bonneau, R., & Ghedin, E. (2018). Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome*, *6*, 12. https://doi.org/10.1186/s40168-017-0393-0

Toju, H., Guimarães, P. R., Olesen, J. M., & Thompson, J. N. (2015). Below-ground plant–fungus network topology is not congruent with above-ground plant–animal network topology. *Scientific Advances*, *1*, e1500291. https://doi.org/10.1126/sciadv.1500291

Tsirogiannis, C., & Sandel, B. (2016). PhyloMeasures: A package for computing phylogenetic biodiversity measures and their statistical moments. *Ecography*, *39*, 709–714. https://doi.org/10.1111/ecog.01814

Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., Grenyer, R., Helmus, M. R., Jin, L. S., Mooers, A. O., Pavoine, S., Purschke, O., Redding, D. W., Rosauer, D. F., Winter, M., & Mazel, F. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology. *Biological Reviews*, *92*, 698–715. https://doi.org/10.1111/brv.12252

Turon, X., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: Separating the wheat from the chaff. *Ecological Applications*, *30*, e02036. https://doi.org/10.1002/eap.2036

U'Ren, J. M., Lutzoni, F., Miadlikowska, J., Zimmerman, N. B., Carbone, I., May, G., & Arnold, A. E. (2019). Host availability drives distributions of fungal endophytes in the imperilled boreal realm. *Nature Ecology & Evolution*, *3*, 1430–1437. https://doi.org/10.1038/s41559-019-0975-2

U'Ren, J. M., Riddle, J. M., Monacell, J. T., Carbone, I., Miadlikowska, J., & Arnold, A. E. (2014). Tissue storage and primer selection influence pyrosequencing-based inferences of diversity and community composition of endolichenic and endophytic fungi. *Molecular Ecology Resources*, *14*, 1032–1048. https://doi.org/10.1111/1755-0998.12252

vanDijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2019). The third revolution in sequencing technology. *Trends in Genetics*, *34*, 666–681. https://doi.org/10.1016/j.tig.2018.05.008

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

Vestheim, H., Deagle, B. E., & Jarman, S. N. (2011). Application of blocking oligonucleotides to improve signal-to-noise ratio in a PCR. *Methods in Molecular Biology*, *687*, 265–274. https://doi.org/10.1007/978-1-60761-944-4_19

Větrovský, T., Baldrian, P., Morais, D., & Berger, B. (2018). SEED 2: A user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics*, *34*, 2292–2294. https://doi.org/10.1093/bioinformatics/bty071

Větrovský, T., Morais, D., Kohout, P., Lepinay, C., Algora, C., Awokunle Hollá, S., Bahnmann, B. D., Bílohnědá, K., Brabcová, V., D'Alò, F., Human, Z. R., Jomura, M., Kolařík, M., Kvasničková, J., Lladó, S., López-Mondéjar, R., Martinović, T., Mašínová, T., Meszárošová, L., … Baldrian, P. (2020). GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. *Scientific Data*, *7*, 228. https://doi.org/10.1038/s41597-020-0567-7

Wagner, A. O., Malin, C., Knapp, B. A., & Illmer, P. (2008). Removal of free extracellular DNA from environmental samples by ethidium monoazide and propidium monoazide. *Applied and Environmental Microbiology*, *74*, 2537–2539. https://doi.org/10.1128/AEM.02288-07

Wang, F., Che, R., Deng, Y., Wu, Y., Tang, L. I., Xu, Z., Wang, W., Liu, H., & Cui, X. (2021). Air-drying and long time preservation of soil do not significantly impact microbial community composition and structure. *Soil Biology and Biochemistry*, *157*, 108238https://doi.org/10.1016/j.soilbio.2021.108238

Wang, H., Qi, J., Xiao, D., Wang, Z., & Tian, K. (2017). A re-evaluation of dilution for eliminating PCR inhibition in soil DNA samples. *Soil Biology and Biochemistry*, *106*, 109–118. https://doi.org/10.1016/j.soilbio.2016.12.011

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*, 5261–5267. https://doi.org/10.1128/aem.00062-07

Washburne, A. D., Morton, J. T., Sanders, J., McDonald, D., Zhu, Q., Oliverio, A. M., & Knight, R. (2018). Methods for phylogenetic analysis of microbiome data. *Nature Microbiology*, *3*, 652–661. https://doi.org/10.1038/s41564-018-0156-0

Webb, C. O., Ackerly, D. D., & Kembel, S. W. (2008). *Phylocom: Software for the Analysis of Community Phylogenetic Structure and Trait Evolution*, *Version 4.0.1* [software]. Retrieved from http://www.phylodiversity.net/phylocom/

Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y. E., Xia, L. C., Xu, Z. Z., Ursell, L., Alm, E. J., Birmingham, A., Cram, J. A., Fuhrman, J. A., Raes, J., Sun, F., Zhou, J., & Knight, R. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME Journal*, *10*, 1669–1681. https://doi.org/10.1038/ismej.2015.235

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., & Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, *5*, 27.

White, J. R., Maddox, C., White, O., Angiuoli, S. V., & Fricke, W. F. (2013). CloVR-ITS: Automated internal transcribed spacer amplicon sequence analysis pipeline for the characterization of fungal microbiota. *Microbiome*, *1*, 6. https://doi.org/10.1186/2049-2618-1-6

White, T. J., Bruns, T. D., Lee, S. B., & Taylor, J. W. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A.Innis, D. H.Gelfand, J. J.Sninsky, & T. J.White (Eds.), *PCR protocols: A guide to methods and applications* (pp. 315–322). Academic Press.

Wickham, H. (2011). Ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*, 180–185. https://doi.org/10.1002/wics.147

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

Wratten, L., Wilm, A., & Göke, J. (2021). Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers.

Nature Methods, *18*, 1161–1168. https://doi.org/10.1038/s41592-021-01254-9

Yang, T., Adams, J. M., Shi, Y., He, J.-S., Jing, X., Chen, L., Tedersoo, L., & Chu, H. (2017). Plant diversity and productivity drive soil fungal richness in natural grasslands of the Tibetan Plateau. *New Phytologist*, *215*, 756–765. https://doi.org/10.1111/nph.14606

Yang, T., Tedersoo, L., Soltis, P. S., Soltis, D. E., Gilbert, J. A., Sun, M., Shi, Y. U., Wang, H., Li, Y., Zhang, J., Chen, Z., Lin, H., Zhao, Y., Fu, C., & Chu, H. (2019). Phylogenetic imprint of woody plants on the soil mycobiome in natural mountain forests of eastern China. *The ISME Journal*, *13*, 686–697. https://doi.org/10.1038/s41396-018-0303-x

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., Gilbert, J. A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., … Glöckner, F. O. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology*, *29*, 415–420. https://doi.org/10.1038/nbt.1823

Yoon, G., Gaynanova, I., & Müller, C. L. (2019). Microbial networks in SPRING - semi-parametric rank-based correlation and partial correlation estimation for quantitative microbiome data. *Frontiers in Genetics*, *10*, 516. https://doi.org/10.3389/fgene.2019.00516

Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., Pavloudi, C., & Pafilis, E. (2020). PEMA: A flexible Pipeline for Environmental DNA Metabarcoding Analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes. *GigaScience*, *9*, giaa022. https://doi.org/10.1093/gigascience/giaa022

Zaiko, A., Greenfield, P., Abbott, C., Ammon, U., Bilewitch, J., Bunce, M., Cristescu, M. E., Chariton, A., Dowle, E., Geller, J., Ardura Gutierrez, A., Hajibabaei, M., Haggard, E., Inglis, G. J., Lavery, S. D., Samuiloviene, A., Simpson, T., Stat, M., Stephenson, S., … Pochon, X. (2022). Towards reproducible metabarcoding data-lessons from an international cross-laboratory experiment. *Molecular Ecology Resources*, *22*, 519–538. https://doi.org/10.1111/1755-0998.13485

Zanne, A. E., Abarenkov, K., Afkhami, M. E., Aguilar-Trigueros, C. A., Bates, S., Bhatnagar, J. M., Busby, P. E., Christian, N., Cornwell, W. K., Crowther, T. W., Flores-Moreno, H., Floudas, D., Gazis, R., Hibbett, D., Kennedy, P., Lindner, D. L., Maynard, D. S., Milo, A. M., Nilsson, R. H., … Treseder, K. K. (2020). Fungal functional ecology: Bringing a trait-based approach to plant-associated fungi. *Biological Reviews*, *95*, 409–433. https://doi.org/10.1111/brv.12570

Zhang, Z. F., Pan, Y. P., Liu, Y., & Li, M. (2021). High diversity of basal fungal lineages and stochastic processes controlled fungal community assembly in mangrove sediments. *Applied and Environmental Microbiology*, *87*, e00928–e1021. https://doi.org/10.1128/AEM.00928-21

Zhou, J., Deng, Y. E., Shen, L., Wen, C., Yan, Q., Ning, D., Qin, Y., Xue, K., Wu, L., He, Z., Voordeckers, J. W., Nostrand, J. D. V., Buzzard, V., Michaletz, S. T., Enquist, B. J., Weiser, M. D., Kaspari, M., Waide, R., Yang, Y., & Brown, J. H. (2016). Temperature mediates continental-scale diversity of microbes in forest soils. *Nature Communications*, *7*, 12083. https://doi.org/10.1038/ncomms12083

Žifčáková, L., Větrovský, T., Howe, A., & Baldrian, P. (2016). Microbial activity in forest soil reflects the changes in ecosystem properties between summer and winter. *Environmental Microbiology*, *18*, 288–301. https://doi.org/10.1111/1462-2920.13026

Zimmerman, N. B., & Vitousek, P. M. (2012). Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proceedings of the National Academy of Sciences*, *109*, 13022–13027. https://doi.org/10.1073/pnas.1209872109

Zinger, L., Bonin, A., Alsos, I. G., Bálint, M., Bik, H., Boyer, F., Chariton, A. A., Creer, S., Coissac, E., Deagle, B. E., De Barba, M., Dickie, I. A., Dumbrell, A. J., Ficetola, G. F., Fierer, N., Fumagalli, L., Gilbert, M. T. P., Jarman, S., Jumpponen, A., … Taberlet, P. (2019). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions. *Molecular Ecology*, *28*, 1857–1862. https://doi.org/10.1111/mec.15060

Zinger, L., Lionnet, C., Benoiston, A. S., Donald, J., Mercier, C., & Boyer, F. (2021). metabaR: An R package for the evaluation and improvement of DNA metabarcoding data quality. *Methods in Ecology and Evolution*, *12*, 586–592. https://doi.org/10.1111/2041-210X.13552

Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., Gaucher, P., Gielly, L., Giguet-Covex, C., Iribar, A., Réjou-Méchain, M., Rayé, G., Rioux, D., Schilling, V., Tymen, B., Viers, J., Zouiten, C., Thuiller, W., Coissac, E., & Chave, J. (2019). Body size determines soil community assembly in a tropical forest. *Molecular Ecology*, *28*, 528–543. https://doi.org/10.1111/mec.14919

Zizka, V. M., Weiss, M., & Leese, F. (2020). Can metabarcoding resolve intraspecific genetic diversity changes to environmental stressors? A test case using river macrozoobenthos. *Metabarcoding and Metagenomics*, *4*, e51925. https://doi.org/10.3897/mbmg.4.51925

SUPPORTING INFORMATION

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

How to cite this article: Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R. H., Kennedy, P. G., Yang, T., Anslan, S., & Mikryukov, V. (2022). Best practices in metabarcoding of fungi: From experimental design to results. *Molecular Ecology*, *31*, 2769–2795. https://doi.org/10.1111/mec.16460