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

Leho Tedersoo[1], Mohammad Bahram[1], Lucie Zinger[2], Henrik Nilsson[3], Peter Kennedy[4], Teng Yang[5], Sten Anslan[6], and Vladimir Mikryukov[7]

[1]University of Tartu
[2]Institut de Biologie de l'ENS (IBENS), Département de Biologie, École normale supérieure, CNRS, INSERM, Université PSL, Paris, France
[3]University of Gothenburg
[4]University of Minnesota
[5]Chinese Academy of Sciences
[6]Institution of Ecology and Earth Sciences
[7]Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences

October 15, 2021

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 about current best practices in metabarcoding analysis of fungal communities, from experimental design through molecular and computational analyses. By re-analysing published datasets, we find that operational taxonomic units (OTUs) outperform amplified sequence variants (ASVs) in recovering fungal diversity, which 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 with the ITS2 subregion. We conclude that metabarcoding analyses of fungi are especially promising for co-analyses with the functional metagenomic or transcriptomic data, integrating fungi in the entire microbiome, recovery of novel fungal lineages and ancient organisms as well as barcoding of old specimens including type material.

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 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 larger fungal clades as well as the kingdom Fungi at large (Gherbawy & Voitk, 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 appropriately 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 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 number of reads 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 (Tedersoo et al., 2014; Sun et al., 2021). 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 address 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 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., 2021a). These long-read technologies and synthetic long reads provide high-quality sequence data for up to 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 et al., 2021a).

Along with the rapid development of HTS methods, bioinformatics platforms and analytical resources have evolved to match the computational needs imposed by large datasets. 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 (Nilsson et al., 2018), eukaryotes more broadly (Ruppert et al., 2019) as well as overall experimental planning (Zinger et al., 2019a), trade-offs among technology generations (Kennedy et al., 2018; Loit et al., 2019), and analytical pitfalls (Halwachs et al., 2017; Critescu & Hebert, 2018). 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 to assess their suitability for fungi. The vast majority of our recommendations are relevant to prokaryotes, protists and metazoans alike.

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 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 et al., 2019a). 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 also require an optimal number of local, biological replicates, ranging from at least three (for detecting strong differences in composition) to around 10 (for richness assessment). 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.

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 pre-treatment 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 at least 5-10 m apart from each other in a homogeneous environment. This distance corresponds to the spatial autocorrelation range in soil fungi (Bahram et al.,

2013) and the average size of 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), although a stratified block design may 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 molds. 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 et al., 2020a).

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) or flame but not 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 et al., 2019a). 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 good quality DNA, 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 an ultra-cold temperature (-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 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 molds (Clasen et al., 2020). Rapid drying methods such as freeze drying and cabinet drying are alternatives to freezing to prevent DNA degradation (Castano et al., 2016). Rapid drying can also be conducted under e.g., the windscreen of a car (at < 40 °C) if a laboratory is inaccessible. 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. Conversely, and 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., 2021). 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, re-analyses for other research purposes or simply for a re-analysis 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.

Molecular analyses

DNA extraction

Prior to DNA extraction, it is important to homogenize the material by using a mortar and pestle, micropestles or bead beating in microcentrifuge tubes. In this process, keeping samples at a relatively cool temperature by using ice cubes or liquid nitrogen is also important. The required amount of material should be weighted to the DNA extraction tube and the rest could be stored for backup or for e.g., 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, plant-debris rich sediments and fleshy plant tissue) may absorb the liquid or co-extract inhibitors. For well-homogenized soil samples, there are only minor differences in richness when using DNA extracts from 0.25 g, 1 g or 10 g material (Song et al., 2015), but increasing the volume (replicate extractions or 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 lab to avoid cross-contamination and air contamination by amplicons. Such potential contaminants can also be detected and removed in downstream analyses through analysis of DNA extraction blank controls.

For DNA extraction, we recommend to follow 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 kits 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 greater with analogous non-robot kits. As a rule of thumb, commercial non-robot and robot-based kits are roughly 2 and 5 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 hamper PCR amplification. These can be overcome by pretreatment of samples during DNA extraction (e.g., Al^{3+} or Ca^{2+} flocculation of humic substances), 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), which 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 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.

Control samples

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 et al., 2019a). Negative and positive controls all inform about external and cross-contamination as well as potential index-switching (Carlsen et al., 2012, Esling et al., 2015). Amplification and sequencing of mock community analysis provide additional insights into the qualitative (i.e., estimation of PCR/sequencing error rates) and quantitative capacity (i.e., biased amplification) to retrieve 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 analyzed, 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 non-biological mock community, containing multiple "species" that are synthetically constructed to have properties equivalent to biological species but never that 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 and index-switching (Ihrmark et al., 2012; Bakker, 2018). 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 processing.

Genetic markers

Obtaining high-quality amplicons is one of the most important steps in metabarcoding analyses. This can be ensured 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-specific and universal primers (Schoch et al., 2012; Nilsson et al., 2018). The ITS region is unsuited to target certain fungi such as Microsporidia (intracellular animal parasites) that may lack this fragment and certain Tulasnellaceae (orchid root symbionts) that have mutations in primer sites (Tedersoo et al., 2015, Rammitsu et al., 2021). 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 (O'Donnell et al., 2015; Cai & Druzhinina, 2021). 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 (Kolarikova et al., 2021). Because of sequencing read length limits imposed by second-generation HTS platforms, researchers have mostly focused on 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 et al., 2015; Tedersoo et al., 2021a).

For metabarcoding, ecologists use mostly primers designed decades ago for Sanger sequencing analyses (Figure 1). 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 variants match to templates. The broadly used fungus-specific forward primer ITS1F is particularly problematic because of several critical mismatches in certain groups of molds 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 containing this intron (Figure 1). It may be important to pair primers with similar melting temperatures to secure 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 ca 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., 2021), 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 4 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* using, e.g. EcoPCR (Ficetola et al., 2010). Indexing both primers with unique tags is more expensive, but allows users to greatly reduce index-switching artefacts (Schnell et al., 2015), and is therefore strongly recommended.

Polymerases

With respect to DNA polymerases, it is important to select one with proofreading capacity in spite of their much greater cost. Proofreading polymerases have much-reduced error rates and therefore result in fewer spurious OTUs (Oliver et al., 2015, Bakker, 2018). 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 and 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.

Thermal cycling conditions

Regarding PCR cycling conditions, reducing annealing temperature may promote amplification of targeted taxa that have one or more primer-template mismatches, but it may also enhance non-specific 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 (Lindahl et al., 2013; D'Amore et al., 2016). 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 BSA may be useful for improving amplification success, but this process may distort the retrieved community (Zaiko et al., 2021).

Biological samples may differ in 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 non-fungal, and the fungal fraction may vary significantly across samples (Tedersoo et al., 2015; Bahram et al., 2018). 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 (Lindahl et al., 2013; Alberdi et al., 2018), but this pooling step will prevent evaluation of the PCR replication and exclusion of dysfunctional PCRs (Taberlet et al., 2018).

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 can be performed 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 ONT (Oxford Nanopore Technology; 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 such as various methylations, it may be possible to record various artificial nucleotide analogues (e.g., 3-bromo-deoxyuridine) incorporated into DNA in real time (Hanson, Allison, Bradford, Wallenstein, & Treseder, 2008; Georgieva et al., 2020). Stable isotope probing is widely used for bacteria because of their rapid metabolism of 13C-enriched substrates (Berry & Loy, 2018), but they have been little used in mycology, likely due to the high costs of enriched C (but see Hannula et al., 2017; Lopez-Mondejar et al., 2020). Nevertheless, RNA-based SIP applications may offer more promise in fungi than for bacteria (Singer et al., 2017; Ghori et al., 2019).

Metagenomics and metatranscriptomics can alternatively 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 within the ongoing Earth Biogenome project (www.earthbiogenome.org) and use of taxonomically more informative long reads. It may also be possible to use targeted capture for rRNA genes or other taxonomically and functional genes to sequence these using long-read protocols (Witek et al., 2016), but the analytical costs are approximately five-fold the costs of regular metabarcoding.

As an alternative to taxon-specific primers, it is possible to use blocking protein-nucleic acid complexes (PNAs) or locked nucleic acid (LNA) oligonucleotides in conjunction with universal PCR primers (Vestheim et al., 2011). PNAs are widely used in metabarcoding analyses of plant-associated bacteria to block amplification and disable subsequent sequencing of plastid and mitochondrial DNA (Lundberg et al., 2013). Probably partly because of primer sites at the end of 18S rRNA gene that allow discrimination against plant amplicons, blocking elements have found limited use in metabarcoding of fungi (but see Ikenaga et al., 2016), although plant-specific motifs exist in all of the 18S, 5.8S, and 28S rRNA genes. Banos et al. (2018) developed protist-targeting PNAs for fungal communities in aquatic environments. The use of blocking elements requires optimisation of concentration and annealing temperature for each primer pair and polymerase used (Vestheim et al., 2011) and furthermore necessitates double-checking any shifts in the perceived diversity of fungi or other target organisms.

DNA library preparation

Among-sample variability of amplicon quantity is high at a low number of PCR cycles. Therefore, the amount of amplicons should be standardised for improved comparability of sequencing depths. This can be achieved by DNA capture on a solid phase with limited binding capacity (SequalPrep, Thermo Fisher Technologies), DNA content measurement and normalisation, or simple estimates of the band strength on agarose gel by eye.

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). In-house library preparation may be up to 5-fold cheaper compared to commercial services.

Sequencing platforms

For metabarcoding, both the second-generation and third-generation platforms can be considered (reviewed in Tedersoo et al., 2021a). Currently, the second-generation platforms allow sequencing up to ca. 550 base

pair 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 x 250 paired-end mode and MiSeq) and MGI-Tech (DNBSEQ-G400RS in 2 x 200 paired-end mode) are best suited for analyses of short barcodes such as ITS1, ITS2 or one or two variable regions combined within 18S and 28S rRNA genes.

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) and 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 markers at high quality. Such long reads offer much improved taxonomic resolution and allow rigorous phylogenetic analyses based on reasonably long alignments of conserved regions (Tedersoo et al., 2020b). 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). 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 et al., 2021a).

Bioinformatics data analysis

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 bioinformatics platforms have been developed for bacterial 16S data, these differ greatly in their capacity of handling fungal ITS sequences that 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 the paired-end 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.

When it comes to quality-filtering, a major step is simultaneous demultiplexing, trimming primer and index sequences and removal of low-quality and non-target 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 incomplete sequence data, respectively. The expected length of full-length ITS ranges from 250 (some*Saccharomycetales*) to around 1,500 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 should not be used as an indicator in quality-filtering for ITS sequences and other non-coding regions.

For ITS metabarcoding, it is important to remove 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 fragments (e.g., 10-70 bases). ITS extraction can be performed using programs cutadapt (Martin et al., 2011), ITSx (Bengtsson-Palme et al., 2013) or ITSxpress (Rivers et al., 2018). Cutadapt is relatively more universal, as it recognizes custom oligonucleotides and cuts from a user-defined distance. ITSx and ITSxpress cut out ITS1, ITS2 and full-length ITS region based on kingdom-wide Hidden Markov Models (HMMs).

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. 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 ca. half of the *de novo* chimeras may be false positives, as suggested for mock communities (Aas et al., 2017). 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. Chimeric molecules usually have only a partial match to the reference sequence (coverage 55-98%) at high sequence similarity (>90%), which allows additional algorithm-based removal or rare, potentially artifactual OTUs.

Index-switches (also known as tag-switches, index-jumping and index cross-talk) is the most deleterious phenomenon in HTS data, because it results in technical cross-contamination among samples and may blur especially patterns in host specificity, taxon networks and biogeographical pattern (Carlsen et al., 2012). Index switches occur during PCR, T4 blunt ending and cross-pairing of amplicons from different libraries, and they are known from all sequencing platforms (Schnell et al., 2005; Carøe & Bohmann, 2020). 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, 2017), the unspread python script (Larsson et al., 2018), by tracking non-biological 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 in vice versa, index-switch rates can be quantified. Sequence abundances above 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) to assess the proportion of index-switch artefacts.

Clustering

Clustering is used for aggregating reads into OTUs based on user-defined sequence similarity thresholds, usually ranging from 95-100% sequence similarity. There are multiple clustering algorithms, many of which 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 similarity (e.g., swarm; Mahe et al., 2021) 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.

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 greedy clustering is recommended as it provides the most stable OTUs by accounting for the taxonomic structure in the reference data (He et al., 2015; Cline et al., 2017).

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 is a better compromise between sequencing errors and biological differences among closely related species and intraspecific variation (Tedersoo et al., 2014; Kyaschenko et al., 2017). 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. However, unicellular fungal lineages and *Glomeromycota* display greater variation, although species-level distinction in the latter group is not straightforward as most virtual taxa and OTUs 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, amplicon sequence variant (ASV) approaches have been developed (Box 2).

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 artifactual insertions, untrimmed fragments of flanking genes or represent an unrecognized 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 BLASTn 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. 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 BLASTn, programs performing k-mer search such as Naïve Bayesian classifier (Wang et al., 2007; Porras-Alfaro et al., 2014) and SINTAX (Edgar, 2016) are up to 100-fold faster, but the resulting similarity estimates are not as straightforward to interpret. These algorithms work well in situations where reference data are abundant and accurately identified to species level, which is a somewhat atypical situation seen to the fungal kingdom at large (Lücking et al., 2021). Phylogenetic placement algorithms such as EPA-ng (Barbera et al., 2018) 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. Some pipelines (e.g., amptk) 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 best hit but also accounting for 4-9 next best hits. Based on multiple global datasets, we have developed recommended taxon-specific e-value and sequence similarity thresholds for 18S-V9, ITS2 and full-ITS reads at the level of genera to phyla (Tedersoo et al., 2014; 2021b; updated in Table S1). Manual BLAST-based double-checking of taxonomic affiliations of at least the, say, 50 largest OTUs/ASVs are a part of those recommendations.

With respect to reference databases, UNITE is the largest by containing curated data obtained from 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 dataset is the so-called 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 non-fungal 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). For *Glomeromycota* 18S and 28S rRNA gene reads, Maarj*AM* (Öpik et al., 2010) and the AM-LSU pipeline (Delavaux et al., 2021) can

be used, respectively.

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. FunFun includes a large number of genomic and enzymatic functional traits, but its taxonomic coverage is limited (Zanne et al., 2020).

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 or 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). Yet these ranks, particularly orders that are generally well-delimited, are informative for understanding the taxonomic 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 analyze the associated primary literature for those taxa and try to determine which functional type is most likely represented based on the study system or research question being addressed.

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 (Froslev et al., 2017). This program allows the removal of OTUs that potentially represent minority haplotypes of common OTUs and remaining 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 by manual browsing, because the first hit may be incompletely annotated. Often, representative sequences require BLASTn-based re-analysis against INSDc for taxonomic determination or chimera control (Nilsson et al., 2012) or more complete functional annotation (Fernandez et al., 2017). Manual BLASTn examination of the 25-50 largest (most frequent or abundant) OTUs is a good way to both identify compromised sequence data and to refine the taxonomic annotation of OTUs. Implementing taxon-specific thresholds for genera to phyla is easier done in a manual mode.

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 require 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 and 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 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 removing

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. It should be borne in mind that sequencing depth per sample is not necessarily related to DNA content of the original sample or biological abundance of certain taxa. The main issue with rarefaction is the substantial loss of data that often corresponds to 90% of the sequencing depth that may carry important information about diversity (McMurdie & Holmes, 2014). Figure 3 illustrates that rarefaction provides relatively lower statistical power in the multivariate analyses as well. Another possibility is rarefaction to median sequencing depth (de Carcer et al., 2011) and accounting for sparse samples in statistical models, or through modelling only (Weiss et al., 2017). An alternative to rarefying is scaling with ranked subsampling, which retains around 20% more OTUs (Beule & Karlovsky, 2020). Some technical methodological comparisons and non-covariate models may still require rarefying.

Statistical data analyses

Alpha diversity metrics

Microbial diversity has been traditionally studied with OTU richness as well as Shannon and Simpson diversity indices and the effective species numbers (Hill numbers) derived from these (Chao et al., 2014; Alberdi & Gilbert, 2019). Diversity indices down-weigh 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. Although commonly used, we do not recommend OTU richness extrapolations, because these rely on the abundance of the rarest OTUs, which are commonly artefactual (Bunge et al., 2014; Balint et al., 2016).

Adding phylogenetic information to taxonomic composition eliminates the uncertainty regarding OTU calculations (Washburne et al., 2018) and could also reduce the effect of remaining PCR/sequencing errors in the data (Taberlet et al 2018). However, phylogenetic composition of fungal communities has rarely been assessed (Horn et al., 2014), because the ITS region is not amenable to robust multiple alignments and phylogenetic reconstruction much beyond the genus level. There are alternative methods available such as grafting phylogenies (Fouquier et al., 2016) and using taxonomic ranks (Tedersoo et al., 2018, Chalmandrier et al., 2019), which may offer some additional insights. Testing phylogenetic conservatism, overdispersion and turnover across phylogenetic scales (Tucker et al., 2016) may also 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 statistics can be calculated in phylocom (Webb et al., 2008), the R packages picante (Kembel et al., 2010) and S.phylomaker (Qian & Jin, 2016) as well as other open-access scripts (Stegen et al., 2015; Ning et al., 2019; Chalmandrier et al., 2019).

Functional data, derived from genomics, metagenomics, metatranscriptomics or functional annotation of OTUs complements information about the taxonomic composition (Laliberte & Legendre, 2010; Egidi et al., 2019; Escalas et al., 2019) because of widespread redundancy and mostly family-level phylogenetic autocorrelation in fungal functional profiles (Zanne et al., 2020). In functional diversity measures, it is important to consider within-function diversity rather than richness or relative abundance in a certain function (Escalas et al., 2019).

Statistical methods

HTS analyses produce semiquantitative abundance data for OTUs, and many rare taxa remain beyond the detection level. Fungal species differ greatly in the number of cells per unit biomass, the number of nuclei per cell and the number of ITS copies per nucleus (Lofgren et al., 2019). The use of quantitative methods assumes that all samples are influenced by these biases in a similar manner.

Because the sequencing data are compositional and reflect relative rather than absolute abundances (Gloor et al., 2017; Lin & Peddada, 2020b), it may be necessary to transform the entire OTU table accordingly for testing factors affecting community shifts (this excludes null model based approaches, e.g. as in Stegen et al.,

2015, which often rely on a sampling process of reads/individuals). Such transformations can be performed by using additive or modified centered log-ratio transformation (Quinn et al., 2019; Yoon et al., 2019; Ha et al., 2020), although it should be kept in mind that these transformations have limits when the data are sparse (Lovell et al., 2020). Alternatively, programs for compositional data analyses (e.g., ANCOM-BC; Lin & Peddada, 2020a) and the Aitchison distance metric can be used (Quinn et al., 2019). With appropriate transformations, different methods reveal roughly comparable results (Figure 3).

In large datasets, our experience is that log-transformation of fungal OTU richness yields better-explained statistical models. 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 also increases the independence of measurements of various ratios summing up 100%.

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 and temporal and plant phylogeny effects on fungal diversity (Zimmerman & Vitousek, 2012; Tedersoo et al., 2020a).

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 Buttigieg & Ramette, 2014; Balint et al., 2016; 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.

When it comes to large amounts of metadata with potential multicollinearity, it is feasible to pre-select the most important variables by machine learning algorithms (Qu et al., 2019) such as random forest implemented in the randomforest (Liaw and Wiener, 2002) and VSURF (Genuer et al., 2019) R packages with subsequent model selection based on forward selection or the AICc information criterion (e.g., nlme R package; Pinheiro et al., 2011; mumin R package). The party package allows estimating interactions among two and 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.

For taxon-level analyses such as distribution modelling of OTUs, 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). Another option is to decompose multivariate ANOVA models into OTU contributions, but this works properly only for unifactorial analyses (Ricotta et al., 2021).

For multivariate analyses, permutational multivariate ANOVA (permanova) is currently the state of the art in that it outperforms common ordination methods because of its explicit hypothesis testing (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. With respect to ordination methods, nonmetric multidimensional scaling (NMDS), principal coordinates analysis (PCoA) and redundancy analysis (RDA) are among the most popular choices for two-dimensional visualisation of multivariate patterns (Paliy & Shankar, 2016). In particular, the ordiellipse function in the vegan package supports analysis and display of within-group variance. General dissimilarity modelling (GDM) facilitates the testing of non-linear effects

of multiple variables, which helps the user to identify and understand critical biological thresholds (gdm R package; Manion et al., 2018). Similar multivariate analyses can also be performed for phylogenetic and functional data. All multivariate analyses are sensitive to selection of transformation and distance (dissimilarity) measures (Kuczynski et al., 2010; Buttigieg & Ramette, 2014)). Hellinger transformation (square-root function) and Bray-Curtis dissimilarity are among the best choices for quantitative data (Legendre & Gallagher, 2001). We do not recommend transforming the data to presences and absences to avoid upweighting rare taxa that may be artefacts.

Network analysis has become increasingly popular in microbial ecology (Faust & Raes, 2012; 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 non-specific taxa are distributed (R package bipartite; Dormann et al., 2009). Both approaches also provide information on the modularity, nestedness and connectivity of the studied ecological network. However, both types of networks have been commonly overinterpreted: unipartite networks may yield incorrect implications based on compositional data (Rao et al., 2021), fail to recover negative associations and exhibit a topology that is sensitive to both network inference techniques and the number of samples considered (Weiss et al., 2016, Matchado et al., 2021); occasional presences are commonly misinterpreted as intimate associations in bipartite networks, particularly for the plant and root-associated fungal relationships. From a more theoretical perspective, it has also been shown that spatial co-occurrences poorly reflect ecological interactions (Blanchet et al., 2020). 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. There are many network construction algorithms (Weiss et al., 2016, Matchado et al., 2021) 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., CoNet; Faust et al., 2016). Some methods take into account data compositionality and sparsity, for example SparCC (Friedman & Alm, 2011), SPIEC-EASI (Kurtz et al., 2015) and SPRING (Yoon et al., 2019)). To reduce the number of false positive associations, it is recommended to reduce matrix size (rare taxa) to reach a fill level of ca. 50% (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 (Weiss et al., 2016; Connor et al., 2017). Nonetheless, consistent associations and integral topological parameters usually remain unaffected (Toju et al., 2015; Röttjers & Faust, 2018).

Structural equation modelling (SEM), in particular path analysis, tests the directionality as well as direct and indirect effects among variables (Fan et al., 2016; Collier, 2020). These are important to consider when the explanatory variables affect each other (e.g., vegetation and soil) or there are several related response variables (e.g., richness of different functional guilds; Yang et al., 2021). Nonetheless, the causal relationship identified by SEMs 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 in SEMs as principal components (Antoninka et al., 2009). The program AMOS (www.ibm.com) and the R package piecewiseSEM (Lefcheck, 2016) offer most functionalities needed for such analyses.

It needs to be remembered that in any ecological tests, the most biologically informative information to the readers are the effect size (either Z-score or adjusted R^2) as well as direction and shape of the relationship. Estimates of variance and sample size are also required for conducting meta-analyses. Values of statistical tests and P -values are context-dependent albeit necessary to report 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).

Absolute and relative abundance

Absolute abundance of bacteria and fungi (as biomass) is an important ecological measure, which indicates ecosystem nutrient cycles. In addition, biomass estimates are important for correctly addressing differences in composition (see below). As fungal biomass cannot be directly measured, there are several proxies for its estimation, notably phospholipid fatty acid (PLFA) profiles, ergosterol abundance, qPCR, and metagenome composition associated with total DNA measurements (Baldrian et al., 2013; Bahram et al., 2018; Yang et al., 2020). However, unicellular fungi are poorly represented in these biomass estimates because of different analogous compounds (Weete et al., 2010), primer biases and the general paucity of reference genome data (Tedersoo et al., 2015). Of the available options, qPCR is the simplest and least laborious method, and it is expected to correlate the strongest with HTS data since comparisons are related to the same DNA extracts. All these methods assume that DNA, sterols and PLFAs are extracted from various samples at equal efficiency, but this cannot be guaranteed for different soils, plant species or tissues.

An alternative is to use a spike-in method, which relies on adding artificial molecules or cells of one or preferably more species that are not expected to occur in biological samples, before DNA extraction (Palmer et al., 2018; Rao et al., 2021). For absolute quantification, it is required that the number of rRNA gene copies of the spike-in taxa is known. Index switches may disproportionally elevate the estimates of spike-in abundance in samples with a high DNA content.

Visualisation of results

There are a large number of methods and software tools available for visualising the statistical results, box plots and scatter plots being 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.

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. 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 samples 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 artifactual OTUs based on ultra-long branches or branches with zero length next to a long-branch taxon (Tedersoo et al., 2020b). Large circular phylogenies squeezed into ca. 17 cm page width offers 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 equipped with large amounts of metadata for display in supplementary materials (Letunic & Bork, 2021).

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. UNITE calculates 100%-similarity OTUs for full-length ITS sequences for each biological sample that are used for further reference and incorporated in SHs. Direct submission of representative sequences to INSDc is discouraged, because these poorly annotated and commonly low-quality data hamper further analyses. We also ask researchers to upload their community matrices, metadata and demultiplexing information (if relevant) in public repositories or supplementary materials, which enables other users to perform meta-analyses and databasing (Põlme et al., 2020; Vetrovsky et al., 2020). Figshare (https://figshare.com/) and Dryad (https://datadryad.org/) are among the most popular choices. Similarly, scripts used for analyses should be released to secure reproducibility and potential reuse in other applications. For the sake of clarity and machine-readability, it is best to use standardised MIMARKS and MIxS terminology (Yilmaz et al., 2011).

In the materials and methods section of scientific papers, 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.

Future perspectives and conclusions

Metabarcoding analyses have recently opened new frontiers in many fields of mycology including assessment of functionality, integration into the entire microbiome analyses, detection of novel fungal taxa and ancient organisms and providing sequence data for taxonomically valuable specimens. Metabarcoding analyses using SIP of labelled substrates (Hannula et al., 2017) or coupled with metatranscriptomics (Zifcakova et al., 2016) and proteomics have revealed functionally active fungal organisms and their activity in situ. Parallel studies of fungi, bacteria and protists have revealed antagonistic interactions (Bork et al., 2015; Bahram et al., 2018), the structure of the micro- and mycobiome web (Tipton et al., 2018) and mechanisms of community assembly (Zinger et al., 2019b). HTS-derived data have revealed several groups of previously undescribed (or unsequenced) order- and class-level fungal lineages (Tedersoo et al., 2020b; Zhang et al., 2021). These reads offer material for constructing taxon-specific primers and probes for visualizing the cells (Chambouvet et al., 2019) and help discovery and characterisation of fungal dark matter. Increasing read length and precision 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), but methods distinguishing organisms representing ancient, recently dead and active organisms should be refined (e.g., metabarcoding along a marker length series, RNA vs. DNA approaches and analyses of chemically modified DNA fraction). Similarly, metabarcoding approaches are useful in generating DNA barcodes from century-old fruit-body specimens including valuable type material where extra care is needed to identify and dismiss air-borne contaminants (Forin et al., 2018).

Metabarcoding of fungi is a rapidly advancing research topic, which has benefited from initial methods and development in clinical microbiology and environmental microbiology - disciplines that are an order of magnitude larger. Thus, fungal ecologists have been able to learn from technological breakthroughs and pitfalls and develop unique solutions for analysis of the ITS marker, the present mainstay of fungal taxonomy and ecology. 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 effect.

Acknowledgments

We thank Mengting Maggie Yuan, Jizhong Zhou and Brendan Furneaux for sharing data and Liis Tiirmann for editing the manuscript.

Author contributions

LT, MB and VM re-analysed data. All authors contributed to ideas and writing.

Data availability statement

This review produced no data and no code.

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. doi: 10.1111/1755-0998.12622 Abdala Asbun, A., Besseling, M. A., Balzano, S., van Bleijswijk, 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. 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. doi: 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. doi: 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.doi: 10.1111/1755-0998.13014 Amend, A., Burgaud, G., Cunliffe, M., Edgcomb, V. P., Ettinger, C. L., Gutierrez, M. H., ... Gladfelter, A. S. (2019). Fungi in the marine environment: open questions and unsolved problems. *mBio, 10*, e01189-18. doi: 10.1128/mBio.01189-18 Amir, A., McDonald, D., Navas-Molina, J.A., Kopylova, E., Morton, J. T., Zech Xu, Z., ... Knight, R. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems, 2*, e00191-16. doi: 10.1128/mSystems.00191-16 Anderson, M. J., Gorley, R. N., & Clarke, R. K. (2008).*Permanova+ for Primer: Guide to Software and Statistical Methods.* Plymouth: Primer-E. Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology, 26,* 32-46. doi: 10.1111/j.1442-9993.2001.01070.pp.x 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. doi: 10.1111/1755-0998.12692 Anslan, S., Li, H., Kunzel, S., & Vences, M. (2021). Microbiomes from feces vs. gut in tadpoles: distinct community compositions between substrates and preservation methods. *Salamandra, 57*, 96-104. doi: 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. doi: 10.3897/mycokeys.39.28109 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. doi: 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. doi: 10.1016/j.funeco.2013.10.003 Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., ... Bork P. (2018). Structure and function of the global topsoil microbiome. *Nature, 560*, 233-237. doi: 10.1038/s41586-018-0386-6 Bahram, M., Peay, K. G., & Tedersoo, L. (2015). Local-scale biogeography and spatiotemporal variability in communities of mycorrhizal fungi. *New Phytologist, 205*, 1454–1463. doi: 10.1111/nph.13206 Bakker, M. G. (2018). A fungal mock community control for amplicon sequencing experiments. *Molecular Ecology Resources, 18*, 541-556. doi: 10.1111/1755-0998.12760 Baldrian, P., Vetrovsky, T., Cajthaml, T., Dobiasova, P., Petrankova, M., Snajdr, J., & Eichlerova, I. (2013). Estimation of fungal biomass in forest litter and soil. *Fungal Ecology, 6*, 1-11. doi: 10.1016/j.funeco.2012.10.002 Balint, M., Bahram, M., Eren, A. M., Faust, K., Fuhrman, J. A., Lindahl, B., ... Tedersoo, L. (2016). Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. *FEMS Microbiology Reviews, 40*, 686-700. doi: 10.1093/fem-

sre/fuw017 Balint, M., Pfenninger, M., Grossart, H. P., Taberlet, P., Vellend, M., Leibold, M. A., ... Bowler, D. (2018). Environmental DNA time series in ecology. *Trends in Ecology & Evolution, 33*, 945-957. doi: 10.1016/j.tree.2018.09.003 Banos, S., Lentendu, G., Kopf, A., Wubet, T., Glockner, F. O., & Reich, M. A. (2018). Comprehensive fungi-specific 18S rRNA gene sequence primer toolkit suited for diverse research issues and sequencing platforms.*BMC Microbiology, 18,* 190. doi: 10.1186/s12866-018-1331-4 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. doi: 10.1093/sysbio/syy054 Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., ... 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. doi: 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. doi: 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. doi: 10.7717/peerj.9593 Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P. M., Reith, F., Dennis, P. G., ... Byrne, M. (2016). Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database.*GigaScience, 5*, 21. doi: 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. doi: 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. doi: 10.1038/ismej.2013.102 Bohmann, K., Elbrecht, V., Caroe, C., Bista, I., Leese, F., Bunce, M., ... & Creer, S. (in press). Strategies for sample labelling and library preparation in DNA metabarcoding studies. *Molecular Ecology Resources*. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology, 37*, 852–857. doi: 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. doi: 10.1126/science.aac5605 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. doi: 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. doi: 10.1111/nph.13849 Brandt, M. I., Trouche, B., Quintric, L., Gunther, 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. doi: 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*,826-826. doi: 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. doi: 10.1146/annurev-statistics-022513-115654 Buschmann, T., & Bystrykh, L. V. (2013). Levenshtein error-correcting barcodes for multiplexed DNA sequencing. *BMC Bioinformatics, 14,*272. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1007/s13225-020-00464-4 Calderon-Sanou, I., Munkemuller, 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. doi: 10.1111/jbi.13681 Callahan, B. J., Grinevich, D., Thakur, S., Balam-

otis, M. A., & Yehezkel, T. B. (2021). Ultra-accurate microbial amplicon sequencing with synthetic long reads. *Microbiome, 9*, 130. doi: 10.1186/s40168-021-01072-3 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. doi: 10.1038/nmeth.3869 Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., ... Dougherty, M. K. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic acids research, 47,* e103. doi: 10.1093/nar/gkz569 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics, 10,* 421. doi: 10.1186/1471-2105-10-421 Carlsen, T., Aas, A. B., Lindner, D., Vralstad, 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. doi: 10.1016/j.funeco.2012.06.003 Caroe, C., & Bohmann, K. (2020). Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples.*Molecular Ecology Resources, 20*, 1620-1631. doi: 10.1111/1755-0998.13227 Castano, C., Parlade, J., Pera, J., De Aragon, 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. doi: 10.1007/s00572-016-0714-3 Castle, S. C., Song, Z., Gohl, D. M., Gutknecht, J. L., Rosen, C. J., Sadowsky, M. J., ... Kinkel, L. L. (2018). DNA template dilution impacts amplicon sequencing-based estimates of soil fungal diversity.*Phytobiomes, 2*, 100-107. doi: 10.1094/PBIOMES-09-17-0037-R Chalmandrier, L., Pansu, J., Zinger, L., Boyer, F., Coissac, E., Genin, A., ... Taberlet, P. (2019). Environmental and biotic drivers of soil microbial β-diversity across spatial and phylogenetic scales.*Ecography, 42*, 2144-2156. doi: 10.1111/ecog.04492 Chambouvet, A., Monier, A., Maguire, F., Itoïz, S., del Campo, J., Elies, P., ... Richards, T. A. (2019). Intracellular Infection of Diverse Diatoms by an Evolutionary Distinct Relative of the Fungi.*Current Biology, 29*, 4093-4101. doi: 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. doi: 10.1890/13-0133.1 Chazdon, R. L., Chao, A., Colwell, R. K., Lin, S. Y., Norden, N., Letcher, S. G., ... Arroyo, J. P. (2011). A novel statistical method for classifying habitat generalists and specialists.*Ecology, 92*, 1332-1343. doi: 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. doi: 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. doi: 10.1111/nph.14752 Collier, J. E. (2020). *Applied structural equation modeling using AMOS: Basic to advanced techniques.* New York: Routledge Publishing. Colwell, R. K., Mao, C. X., & Chang, J. (2004). Interpolating, extrapolating, and comparing incidence-based species accumulation curves. *Ecology, 85*, 2717–2727. doi: 10.1890/03-0557 Connor, N., Barberan, A., & Clauset, A. (2017). Using null models to infer microbial co-occurrence networks. *PLoS ONE, 12*, e0176751. doi: 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. doi: 10.1146/annurev-ecolsys-110617-062306 Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data.*Microbiome, 6*, 226. doi: 10.1186/s40168-018-0605-2 de Carcer, 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. doi: 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. doi: 10.1002/ece3.6219 Delavaux, C. S., Sturmer, S. L., Wagner, M. R., Schutte, 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. doi: 10.1111/nph.17080 Deng, Y., Jiang, Y-H., Yang, Y., He, Z., Luo, F., & Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinformatics, 13*, 113. doi: 10.1186/1471-2105-13-113 Deshpande, V., Wang, Q.,

Greenfield, P., Charleston, M., Porras-Alfaro, A., Kuske, C. R., ... Tran-Dinh, N. (2016). Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia, 108*, 1-5. doi: 10.3852/14-293 Dickie, I. A., Boyer, S., Buckley, H. L., Duncan, R. P., Gardner, P. P., Hogg, I. D., ... Weaver, L. (2018). Towards robust and repeatable sampling methods in eDNA-based studies. *Molecular Ecology Resources, 18*, 940–952. doi: 10.1111/1755-0998.12907 Dormann, C. F., Frund, J., Bluthgen, N., & Gruber, B. (2009). Indices, graphs and null models: analyzing bipartite ecological networks.*Open Ecology Journal, 2*, 7-24. doi: 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. doi: 10.1016/j.funbio.2011.03.005 Dray, S., Blanchet, G., Borcard, D., Guenard, G., Jombart, T., Larocque, G., ... Wagner, H. H. (2018). *Package 'adespatial'.*Available:https://cran.microsoft.com/web/packages/adespatial/adespatial.pdf 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. doi: 10.1093/bioinformatics/btr381 Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*, 2460-2461. doi: 10.1093/bioinformatics/btq461 Edgar, R. C. (2016). UNOISE2: Improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv, 2016*, 081257. doi: 10.1101/081257 Edgar, R. C. (2017). Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ, 5*, e3889. doi: 10.7717/peerj.3889 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. doi: 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. doi: 10.1016/j.tim.2018.11.003 Escalas, A., Hale, L., Voordeckers, J. W., Yang, Y., Firestone, M. K., Alvarez-Cohen, L., & Zhou, J. (2019). Microbial functional diversity: From concepts to applications. *Ecology and Evolution, 9*, 12000-12016. doi: 10.1002/ece3.5670 Escudie, F., Auer, L., Bernard, M., Mariadassou, M., Cauquil, L., Vidal, K., ... Pascal, G. (2018). FROGS: Find, rapidly, OTUs with galaxy solution. *Bioinformatics, 34,* 1287-1294. doi: 10.1093/bioinformatics/btx791 Estensmo, E. L., Maurice, S., Morgado, L., Martin-Sanchez, P. M., Skrede, I., & Kauserud, H. (2011). The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species. *Molecular Ecology Resources, 21*, 1141-1148. doi: 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. doi: 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. doi: 10.1186/s13717-016-0063-3 Faust, K., & Raes, J. (2016). CoNet app: Inference of biological association networks using Cytoscape. *F1000Research, 5*, 1519. doi: 10.12688/f1000research.9050.2 Faust, K., & Raes, J. (2012). Microbial interactions: From networks to models. *Nature Reviews Microbiology, 10*, 538-549. doi: 10.1038/nrmicro2832 Fernandez, C. W., Nguyen, N. H., Stefanski, A., Han, Y., Hobbie, S. E., Montgomery, R. A., ... 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. doi: 10.1111/gcb.13510 Ficetola, G. F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessiere, J., ... Pompanon, F. (2010). An in silico approach for the evaluation of DNA barcodes. *BMC Genomics, 11*, 434. doi: 10.1186/1471-2164-11-434 Fischer, M., Renevey, N., Thur, B., Hoffmann, D., Beer, M., & Hoffmann, B. (2016). Efficacy assessment of nucleic acid decontamination reagents used in molecular diagnostic laboratories. *PLoS ONE, 11*, e0159274. doi: 10.1371/journal.pone.0159274 Foster, Z. S. L., Sharpton, T. J., & Grunwald, N. J. (2017). Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Computational Biology, 13*, e1005404. doi: 10.1371/journal.pcbi.1005404 Fouquier, J., Rideout, J. R., Bolyen, E., Chase, J., Shiffer, A., McDonald, D., ... Kelley, S. T. (2016). Ghost-tree: Creating hybrid-gene phylogenetic trees for diversity analyses. *Microbiome, 4*, 11. doi: 10.1186/s40168-016-0153-6 Friedman, J., & Alm, E. J. (2011). Inferring correlation networks from genomic survey data. *PLoS Computational Biology, 8*, e1002687. doi: 10.1371/journal.pcbi.1002687 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. doi: 10.1111/1755-0998.13387 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. doi: 10.1038/ismej.2015.30 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. doi: 10.1093/nar/gkaa517 Gherbawy, Y., & Voigt, K. (eds). (2010). *Molecular identification of fungi.* Berlin: Springer. Glassman, S. I., & Martiny, J. B. (2018). Broad-scale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. *mSphere, 3*, e00148-18. doi: 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. doi: 10.3389/fmicb.2017.02224 Gohl, D. M., Auch, B., Certano, A., LeFrancois, B., Bouevitch, A., Doukhanine, E., ... Beckman, K. B. (2021). Dissecting and tuning primer editing by proofreading polymerases. *Nucleic Acids Research, 49*, e87. doi: 10.1093/nar/gkab471 Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., ... Knights, D. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nature Biotechnology, 34*, 942-949. doi: 10.1038/nbt.3601 Gotelli, N. J., & Ellison, A. M. (2004). *A primer of ecological statistics.* Sunderland: 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. doi: 10.1038/s41579-019-0175-8 Guerrieri, A., Bonin, A., Munkemuller, T., Gielly, L., Thuiller, W., & Ficetola, G. F. (2021). Effects of soil preservation for biodiversity monitoring using environmental DNA. *Molecular Ecology, 30,*3313-3325. doi: 10.1111/mec.15674 Gweon, H. S., Oliver, A., Taylor, J., Booth, T., Gibbs, M., Read, D. S., ... Schonrogge, K. (2015). PIPITS: An automated pipeline for analyses of fungal internal transcribed spacer sequences from the I llumina sequencing platform. *Methods in Ecology and Evolution, 6,*973-980. doi: 10.1111/2041-210x.12399 Ha, M. J., Kim, J., Galloway-Pena, J., Do, K. A., & Peterson, C. B. (2020). Compositional zero-inflated network estimation for microbiome data. *BMC Bioinformatics, 21*, 581. doi: 10.1186/s12859-020-03911-w Halwachs, B., Madhusudhan, N., Krause, R., Nilsson, R. H., Moissl-Eichinger, C., Hogenauer, C., & Gorkiewicz, G. (2017). Critical issues in mycobiota analysis. *Frontiers in microbiology, 8*, 180. doi: 10.3389/fmicb.2017.00180 Hamady, M., Lozupone, C., & Knight, R. (2010). Fast Unifrac: Facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data.*The ISME Journal, 4*, 17-27. doi: 10.1038/ismej.2009.97 Hannula, S. E., Morrien, E., de Hollander, M., van der Putten, W. H., van Veen, J. A., & de Boer, W. (2017). Shifts in rhizosphere fungal community during secondary succession following abandonment from agriculture. *The ISME Journal, 11*, 2294–2304. doi: 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. doi: 10.1007/s10021-008-9186-4 He, Y., Caporaso, J. G., Jiang, X-T., Sheng, H-F., Huse, S. M., Rideout, J. R., ... Zhou, H-W. (2015). Stability of operational taxonomic units: An important but neglected property for analyzing microbial diversity. *Microbiome, 3*, 20. doi: 10.1186/s40168-015-0081-x Heeger, F., Bourne, E. C., Baschien, C., Yurkov, A., Bunk, B., Sproer, C., ... 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. doi: 10.1111/1755-0998.12937 Hildebrand, F., Tadeo, R., Wright, A. Y., Bork, P., & Raes, J. (2014). LotuS: An efficient and user-friendly OTU processing pipeline.*Microbiome, 2*, 30.doi: 10.1186/2049-2618-2-30 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. doi: 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. doi: 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. doi: 10.3389/fmicb.2017.01561 Ihrmark, K., Bodeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenk, J., ... 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. doi: 10.1111/j.1574-6941.2012.01437.x Ikenaga, M., Tabuchi, M., Kawauchi, T., & Sakai, M. (2016). Application of Locked Nucleic Acid (LNA) primer and PCR clamping by LNA oligonucleotide to enhance the amplification of Internal Transcribed Spacer (ITS) regions in investigating the community structures of plant–associated fungi. *Microbes and Environments, 31*, 339-348. doi: 10.1264/jsme2.me16085 Jalili, V., Afgan, E., Gu, Q., Clements, D., Blankenberg, D., Goecks, J., ... Nekrutenko, A. (2020). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update.*Nucleic Acids Research, 48*, W395-W402. doi: 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. doi: 10.1038/s41559-018-0614-3 Joos, L., Beirinckx, S., Haegeman, A., Debode, J., Vandecasteele, B., Baeyen, S., & 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. doi: 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. doi: 10.1111/j.1469-8137.2009.02990.x Karst, S. M., Ziels, R. M., Kirkegaard, R. H., Sorensen, E. A., McDonald, D., Zhu, Q., ... Albertsen, M. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with nanopore or PacBio sequencing. *Nature Methods, 18*, 165-169. doi: 10.1038/s41592-020-01041-y Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., ... Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics, 26*, 1463-1464. doi: 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. doi: 10.1111/nph.14883 Kolarikova, Z., Slavikova, R., Kruger, C., Kruger, 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. doi: 10.1111/nph.17372 Koljalg, U., Nilsson, H. R., Schigel, D., Tedersoo, L., Larsson, K. H., May, T. W., ... Abarenkov, K. (2020). The Taxon Hypothesis paradigm—on the unambiguous detection and communication of taxa.*Microorganisms, 8*, 1910. doi: 10.3390/microorganisms8121910 Koljalg, U., Tedersoo, L., Nilsson, R. H., & Abarenkov, K. (2016). Digital identifiers for fungal species. *Science, 352*, 1182-1183. doi: 10.1126/science.aaf7115 Kuczynski, J., Liu, Z., Lozupone, C., McDonald, D., Fierer, N., & Knight, R. (2010). Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. *Nature Methods, 7*, 769-775. doi: 10.1038/nmeth.1499 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. doi: 10.1111/1755-0998.13107 Kurtz, Z. D., Muller, 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. doi: 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. doi: 10.1038/ismej.2016.184 Lagerborg, K. A., Normandin, E., Bauer, M. R., Adams, G., Figueroa, K., Loreth, C., ... Hooper, D. (2021). DNA spike-ins enable confident interpretation of SARS-CoV-2 genomic data from amplicon-based sequencing. *bioRxiv, 2021*, 435654. doi: 10.1101/2021.03.16.435654 Laliberte, E., & Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology, 91*, 299-305. doi: 10.1890/08-2244.1 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. doi: 10.1038/nmeth.4666 Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia, 129*, 271-280. doi: 10.1007/s004420100716 Legendre, P., & Legendre, L. (2012). *Numerical Ecology*. New York: Elsevier. Lennon, J. T., Muscarella, M. E., Placella, S. A., & Lehmkuhl, B. K. (2018). How, when, and where relic DNA affects microbial diversity.*mBio, 9*, e00637-18. doi: 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. doi: 10.1093/nar/gkab301 Li, S., Du, X., Feng, K., Wu, Y., He, Q., Wang, Z., ... Escalas, A. (2021). Assessment of microbial α-diversity in one meter squared topsoil. *Soil Ecology Letters*, *2021*, 1-3. doi: 10.1007/s42832-021-0111-5 Lin, H., & Peddada, S.

D. (2020a). Analysis of compositions of microbiomes with bias correction. *Nature Communications, 11*, 3514. doi: 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. doi: 10.1038/s41522-020-00160-w Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., ... Kauserud, H. (2013). Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide.*New Phytologist, 199*, 288–299. doi: 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. doi: 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. doi: 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. doi: 10.1111/mec.14995 Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, 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*, e01368-19. doi: 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*, 720513. doi: 10.3389/fmicb.2021.720513 Lopez-Mondejar, R., Tlaskal, V., Větrovský, T., Štursová, M., Toscan, R., da Rocha, 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*, 107875. doi: 10.1016/j.soilbio.2020.107875 Lovell, D. R., Chua, X. Y., & McGrath, A. (2020). Counts: an outstanding challenge for log-ratio analysis of compositional data in the molecular biosciences. *NAR Genomics and Bioinformatics, 2*, lqaa040. Lücking, R., Aime, M. C., Robbertse, B., Miller, A. N., Aoki, T., Ariyawansa, H. A., ... & Schoch, C. L. (2021). Fungal taxonomy and sequence-based nomenclature. *Nature Microbiology, 6*, 540-548. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nature Methods, 10*, 999–1002. doi: 10.1038/nmeth.2634 Mahe, F., Czech, L., Stamatakis, A., Quince, C., de Vargas, C., Dunthorn, M., & Rognes, T. (in press). Swarm v3: Towards tera-scale amplicon clustering. *Bioinformatics*. doi: 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 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet, 17*, 10-12. doi: 10.14806/ej.17.1.200 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 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. doi: 10.1134/S1067413621030085 Miyauchi, S., Kiss, E., Kuo, A., Drula, E., Kohler, A., Sanchez-Garcia, M., ... Buee, M. (2020). Large-scale genome sequencing of mycorrhizal fungi provides insights into the early

evolution of symbiotic traits. *Nature Communications, 11,* 5125. doi: 10.1038/s41467-020-18795-w Moreira-Grez, B., Vuong, P., Waite, I., Morald, T., Wise, M., & Whiteley, A. S. (2019). RNA Stable Isotope Probing (RNA-SIP).*Methods in Molecular Biology, 2046*, 31-44. doi: 10.1007/978-1-4939-9721-3_3 Nguyen, N. H., Smith, D., Peay, K., & Kennedy, P. (2015). Parsing ecological signal from noise in next generation amplicon sequencing.*New Phytologist, 205*, 1389–1393. doi: 10.1111/nph.12923 Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., ... Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community data sets by ecological guild. *Fungal Ecology, 20*, 241–248. doi: 10.1016/j.funeco.2015.06.006 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. doi: 10.1038/s41579-018-0116-y Nilsson, R. H., Larsson, K. H., Taylor, A. F., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., ... Abarenkov, K. (2019). The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research, 47,*D259-D264. doi: 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., ... Koljalg, U. (2012). Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys, 4*, 37–63. doi: 10.3897/mycokeys.4.3606 Noble, W. S. (2009). How does multiple testing correction work?*Nature biotechnology, 27*, 1135-1137. doi: 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. doi: 10.1007/s12600-015-0484-z Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... 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. doi: 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. doi: 10.1186/1471-2105-12-385 Opik, M., Vanatoa, A., Vanatoa, E., Moora, M., Davison, J., Kalwij, J. M., ... Zobel, M. (2010). The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (*Glomeromycota*). *New Phytologist, 188,* 223-241. doi: 10.1111/j.1469-8137.2010.03334.x Paliy, O., & Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology, 25*, 1032–1057. doi: 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. doi: 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. doi: 10.7717/peerj.6876 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. doi: 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. doi: 10.3390/microorganisms9010188 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. doi: 10.1128/aem.02894-13 Polme, S., Abarenkov, K., Nilsson, R. H., Lindahl, B. D., Clemmensen, K. E., Kauserud, H., ... Tedersoo, L. (2020). FungalTraits: A user-friendly traits database of fungi and fungus-like stramenopiles.*Fungal Diversity, 105*, 1–16. doi: 10.1007/s13225-020-00466-2 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. doi: 10.1093/jpe/rtv047 Quast, C.,

Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research, 41*, D590–D596. doi: 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. doi: 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. doi: 10.1093/gigascience/giz107 Radajewski, S., McDonald, I. R., & Murrell, J. C. (2003). Stable-isotope probing of nucleic acids: A window to the function of uncultured microorganisms. *Current Opinion in Biotechnology, 14*, 296-302. doi: 10.1016/S0958-1669(03)00064-8 Rajala, T., Peltoniemi, M., Hantula, J., Makipaa, R., & Pennanen, T. (2011). RNA reveals a succession of active fungi during the decay of Norway spruce logs. *Fungal Ecology, 4*, 437–448. doi: 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. doi: 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. doi: 10.1038/s41586-021-03241-8 Ricotta, C., Pavoine, S., Cerabolini, B. E., & Pillar, V. D. (2021). A new method for indicator species analysis in the framework of multivariate analysis of variance. *Journal of Vegetation Science, 32*, e13013. doi: 10.1111/jvs.13013 Rillig, M. C., Ryo, M., Lehmann, A., Aguilar-Trigueros, C. A., Buchert, S., Wulf, A., ... Yang, G. (2019). The role of multiple global change factors in driving soil functions and microbial biodiversity.*Science, 366*, 886-890. doi: 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. doi: 10.12688/f1000research.15704.1 Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahe, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ, 4,*e2584. doi: 10.7717/peerj.2584 Roy, J., Mazel, F., Sosa-Hernandez, M. A., Duenas, 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. doi: 10.1111/nph.16080 Rottjers, L., & Faust, K. (2018). From hairballs to hypotheses – biological insights from microbial networks. *FEMS Microbiology Reviews, 42*, 761-780. doi: 10.1093/femsre/fuy030 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. doi: 10.1038/s41598-017-14978-6 Sato, M. P., Ogura, Y., Nakamura, K., Nishida, R., Gotoh, Y., Hayashi, M., ... 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. doi: 10.1093/dnares/dsz017 Schloss, P. D. (2020). Reintroducing mothur: 10 years later.*Applied and Environmental Microbiology, 86,* e02343-19. doi: 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. doi: 10.1111/1755-0998.12402 Schwarzenbach, K., Enkerli, J., & Widmer, F. (2007). Objective criteria to assess representativity of soil fungal community profiles.*Journal of Microbiological Methods, 68*, 358-366. doi: 10.1016/j.mimet.2006.09.015 Singer, E., Wagner, M., & Woyke, T. (2017). Capturing the genetic makeup of the active microbiome in situ. *The ISME Journal, 11*, 1949-1963. doi: 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. doi: 10.1371/journal.pone.0127234 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. doi: 10.1128/mSphere.00163-19 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. doi: 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. doi: 10.1101/2020.07.02.185710 Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018).*Environmental DNA: For biodiversity research and monitoring.* Lon-

don: 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. doi: 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. doi: 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. doi: 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. doi: 10.1890/12-1693.1 Tedersoo, L., Albertsen, M., Anslan, S., & Callahan, B. (2021a). Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. *Applied and Environmental Microbiology, 87*, e00626-21. doi: 10.1128/aem.00626-21 Tedersoo, L., Anslan, S., Bahram, M., Drenkhan, R., Pritsch, K., Buegger, F., ... 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. doi: 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. doi: 10.1007/s13225-020-00456-4 Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., ... Abarenkov, K. (2015). Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys, 10*, 1-43. doi: 10.3897/mycokeys.10.4852 Tedersoo, L. et al. 2021b. GSMc. *Fungal Diversity*, pending revision. Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., ... Abarenkov, K. (2014). Global diversity and geography of soil fungi. *Science, 346*, 1078. doi: 10.1126/science.1256688 Tedersoo, L., & Lindahl, B. (2016). Fungal identification biases in microbiome projects. *Environmental Microbiology Reports, 8*, 774-779. doi: 10.1111/1758-2229.12438 Tipton, L., Müller, C. L., Kurtz, Z. D., Huang, L., Kleerup, E., Morris, A., ... Ghedin, E. (2018). Fungi stabilize connectivity in the lung and skin microbial ecosystems. *Microbiome, 6*, 12. doi: 10.1186/s40168-017-0393-0 Tucker, C. M., Cadotte, M. W., Carvalho, S. B., Davies, T. J., Ferrier, S., Fritz, S. A., ... Pavoine, S. (2017). A guide to phylogenetic metrics for conservation, community ecology and macroecology.*Biological Reviews, 92,* 698-715. doi: 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. doi: 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. doi: 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. doi: 10.1111/1755-0998.12252 van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2019). The third revolution in sequencing technology. *Trends in Genetics, 34*, 666-681. doi: 10.1016/j.tig.2018.05.008 Vasar, M., Davison, J., Neuenkamp, L., Sepp, S. K., Young, J. P. W., Moora, M., & Opik, M. (2021). User-friendly bioinformatics pipeline gDAT (graphical downstream analysis tool) for analysing rDNA sequences.*Molecular Ecology Resources, 21*, 1380-1392. 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. doi: 10.1007/978-1-60761-944-4_19 Vetrovsky, 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. doi: 10.1093/bioinformatics/bty071 Vetrovsky, T., Morais, D., Kohout, P., Lepinay, C., Algora, C., Awokunle Holla, S., ... Baldrian, P. (2020). GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. *Scientific Data, 7*, 228. doi: 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. doi: 10.1128/AEM.02288-07 Wang, F., Che, R., Deng, Y., Wu, Y., Tang, L., Xu, Z., ... Cui, X. (2021). Air-drying and long time preservation of soil do not signifi-

cantly impact microbial community composition and structure.*Soil Biology and Biochemistry, 157*, 108238. 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. doi: 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. doi: 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. doi: 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/ Weete, J. D., Abril, M., & Blackwell, M. (2010). Phylogenetic distribution of fungal sterols. *PloS ONE, 5,* e10899. doi: 10.1371/journal.pone.0010899 Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., ... Birmingham, A. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME Journal, 10*, 1669-1681. doi: 10.1038/ismej.2015.235 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. doi: 10.1186/2049-2618-1-6 Witek, K., Jupe, F., Witek, A. I., Baker, D., Clark, M. D., & Jones, J. D. (2016). Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. *Nature Biotechnology, 34,*656-660. doi: 10.1038/nbt.3540 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. doi: 10.1111/nph.14606 Yoon, G., Gaynanova, I., & Muller, 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. doi: 10.3389/fgene.2019.00516 Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., ... 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. doi: 10.1093/gigascience/giaa022 Zaiko, A., Greenfield, P., Abbott, C., von Ammon, U., Bilewitch, J., Bunce, M., Cristescu, M. E., Chariton, A., Dowle, E., Geller, J., & Ardura Gutierrez, A. (in press). Towards reproducible metabarcoding data-lessons from an international cross-laboratory experiment.*Molecular Ecology Resources*. doi: 10.1111/1755-0998.13485 Zanne, A. E., Abarenkov, K., Afkhami, M. E., Aguilar-Trigueros, C. A., Bates, S., Bhatnagar, J. M., ... Moreno, H. F. (2020). Fungal functional ecology: Bringing a trait-based approach to plant-associated fungi. *Biological Reviews, 95*, 409-433. doi: 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,* e00928-21. doi: 10.1128/AEM.00928-21 Zhou, J., Deng, Y., Shen, L., Wen, C., Yan, Q., Ning, D., ... Brown, J. H. (2016). Temperature mediates continental-scale diversity of microbes in forest soils. *Nature Communications, 7*, 12083. doi: 10.1038/ncomms12083 Zifcakova, L., Vetrovsky, 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. doi: 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. doi: 10.1073/pnas.1209872109 Zinger, L., Bonin, A., Alsos, I. G., Balint, M., Bik, H., Boyer, F., ... De Barba, M. (2019a). DNA metabarcoding—Need for robust experimental designs to draw sound ecological conclusions.*Molecular Ecology, 28*, 1857-1862. doi: 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. doi: 10.1111/2041-210X.13552 Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., ... Chave, J. (2019b). Body size determines soil community assembly in a tropical forest. *Molecular Ecology, 28*, 528-543. doi: 10.1111/mec.14919

Table 1. Properties of bioinformatics pipelines used for fungal metabarcoding. Interface: command line (CL).

Software	Interface[1]	Type	Target group	ITS extraction	U
QIIME2	graphical, CL	offline, multiple included dependencies	prokaryote, short-read	ITSxpress	in
mothur	CL	offline	prokaryote, short-read	-	C
Claident	CL	offline	universal, short-read	ITSx	n
PIPITS	CL	offline	fungi, ITS1, ITS2	ITSx	-
SEED2	graphical	offline	universal, short-read	ITSx compatible	re
PEMA	CL	offline, multiple included dependencies	universal, short-read	-	p
PipeCraft	graphical	offline, multiple included dependencies	universal, short/long-read	ITSx	m
DADA2	CL (R-Studio)	offline	universal, short/long-read	-	A
LOTUS2	CL	offline	universal, short/long-read	ITSx	lo
Galaxy	graphical	online	prokaryote, short-read	-	li
FROGS	graphical	online	prokaryote, short-read	ITSx	G
AMPtk	CL	offline	universal, short-read	-	h
OBITools	CL	offline	universal, short-read	-	-
SCATA	graphical	online	fungi, short/long-read	ITSx	u
gDAT	graphical	offline	universal, short-read	-	-
Cascabel	CL	offline	universal, short-read	-	g
MICCA	CL	offline	universal, short-read	-	C
SWARM	CL	offline	universal, short-read	-	lo
CloVR-ITS	CL	offline/online	fungi, short-read	-	V
DAnIEL	graphical	online	fungi, short-read	ITSx (via PIPITS)	ra

[1]CL, command line

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

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.

Figure 3. Non-metric multidimensional scaling (NMDS) graphs illustrating relative performance of various 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) non-rarefied data; (c, d) rarefied data; (e, f) scaling with ranked subsampling (SRS) normalised data; (g, h) centered log-ratio (CLR) transformed data. Numbers on symbols indicate plant species (separate numbering for terrestrial and aquatic plants); ellipses depict 95% CI for tissue and habitat combinations. Explained variation (%) as revealed from Permanova+ analysis is indicated (t x h, tissue and habitat interaction; seqs, sequencing depth). Plant species effects are not analysed here for simplicity. Data from A. Azadnia, V. Mikryukov, L. Tedersoo (unpublished).

Box1. Trade-offs in Sample Pooling

To improve representativeness of the samples at minimum extra cost, pooling statistically non-independent 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 Box1). The number of subsam-

ples 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 estimate 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, e.g. *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), e.g. for regional- to global-scale analyses.

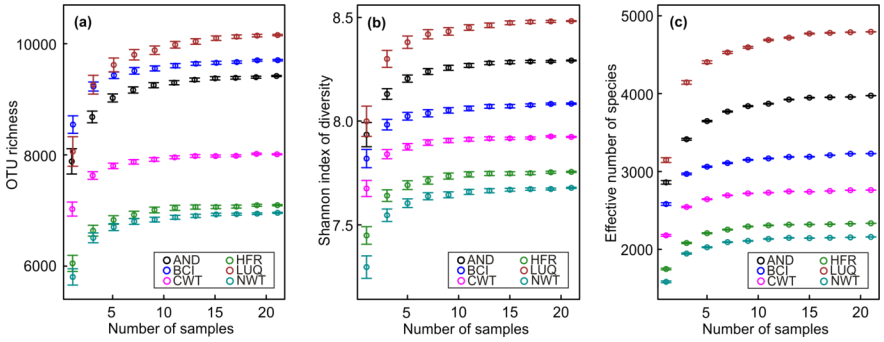

Figure Box1. Potential underestimation of biodiversity and high variance at low number of (sub)samples: (a) OTU richness, (b) Shannon index of diversity and (c) Effective number of species (here as the exponential of the Shannon index). Rarefied datasets from six sites (Zhou et al., 2016) were randomized 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 both rarefaction steps. Symbols represent means of analytically pooled samples, and error bars indicate 95% CI. The differences between a single sample and 21 randomly selected samples average 18.7% (+-3.6%), 4.4% (+-0.9%) and 35.7% (+-7.9%) for (a), (b) and (c), respectively.

Box 2. Amplified sequence variant (ASV) methods

ASV approaches represent a specific type of greedy *de novo*clustering and several alternative methods - notably DADA2 (Callahan et al., 2016), UNOISE (Edgar, 2016) and deblur (Amir et al., 2017) - have recently become popular in microbiology including fungal ecology (Glassman et al., 2018). In DADA2, ASVs correspond to 100%-similarity OTUs, where sample-wise rare variants are assigned to dominant haplotypes based on an error model with stringent settings (Callahan et al., 2016). Deblur is relatively less conservative (Amir et al., 2017) and relatively inefficient in removing rare haplotypes (Li et al., 2021). The ASV approaches are certainly useful for separating as many taxa as possible based on conserved genes, but their utility for ITS and protein-encoding genes is unclear. They may outperform traditional approaches in distinguishing the aforementioned saprotrophic and pathogenic *Ascomycota* with haploid genomes. However, it severely biases diversity estimates of metazoans based on the Cytochrome Oxidase 1 (CO1) gene (Brandt et al., 2021) and

it is expected to perform poorly for fungal groups with dikaryotic (*Basidiomycota*), diploid (most unicellular groups), and polyploid (*Glomeromycota*) genomes that commonly exhibit two or multiple different rRNA gene and ITS copies per genome or even within haploid nuclei (Lindner et al., 2013; Egan et al., 2018). Estensmo et al. (2021) demonstrated that in polypores, the ASVs significantly overestimated species richness. Using a re-analysis of a dataset from Furneaux et al. (2021), we show that the ASV approach reduces phylogenetic richness by disproportionately eliminating rare members of the unicellular fungal groups, *Glomeromycota* and non-fungal eukaryotes (Figure Box2). In terms of beta diversity, the results are similar between ASV and OTU-based approaches (Glassman et al., 2018). If one is interested in using the denoising algorithm of DADA2, a subsequent post-clustering of ASVs (implemented in pipelines such as LotuS2 and amptk) may be a solution (Furneaux et al., 2021; Estensmo et al., 2021), but this approach does not ameliorate the loss of true unique species.

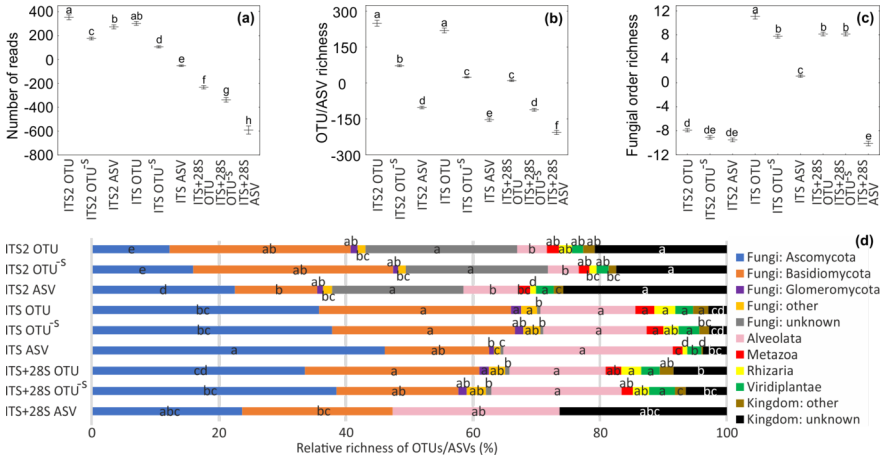

Figure Box2. Comparison of OTU-based (OTU-**s** , singletons removed) and ASV-based approaches for inferring taxonomic richness using the ITS2 subregion, full ITS region, and ITS + 28S based on the dataset of Furneaux et al. (2021), which was taxonomically re-analysed following Tedersoo et al. (in press). (a) average number of reads retained; (b) richness of OTUs/ASVs; (c) richness of fungal orders; and (d) kingdom and phylum-level distribution of OTUs/ASVs across the entire dataset. In (d), letters indicate statistically significant differences among groups based on Scheffe PostHoc tests following sample-wise testing of log-ratio transformed proportions using two-way ANOVAs including the two factors marker length and matrix type - OTU, OTUs without singletons (OTU-s) and ASV. We conclude that 1) the ASV approach and OTU-based approach with singletons removed recover lower proportions of non-Dikarya and non-fungal taxa compared with the OTU-based approach including singletons; 2) analyses based on the ITS2 region alone revealed a higher proportion of OTUs/ASVs that could not be identified to a fungal phylum or eukaryote kingdom compared with longer marker fragments due to less taxonomic information; 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; and 4) the ASV approach performs suboptimally for metabarcoding of fungi and particularly poorly for the full ITS region and longer markers.