Fungal Ecology 65 (2023) 101274

Contents lists available at ScienceDirect

Fungal Ecology

journal homepage: www.elsevier.com/locate/funeco

Mini-review

ITS alchemy: On the use of ITS as a DNA marker in fungal ecology

Håvard Kauserud

Sections for Genetics and Evolutionary Biology, Department of Biosciences, University of Oslo, Norway

ARTICLE INFO

Handling Editor: Dr. Sarah Christofides

Index descriptors:
Internal transcribed spacer
ITS
Fungi
Metabarcoding
Community ecology

ABSTRACT

High throughput sequencing of PCR amplicons derived from environmental DNA (aka DNA metabarcoding) has become an integral part of fungal ecology, enabling in-depth characterization of fungal communities. In most cases, the rDNA Internal Transcribed Spacer (ITS) region, which has a long history as a target in fungal systematics, is used as a DNA barcode marker. Despite improvements in sequencing techniques and bioinformatics approaches, there are inherent limitations associated with the use of a single-locus DNA marker that are often ignored. In this text, I discuss both inherent biological and methodological limitations associated with the use of the ITS marker. For example, proper species delimitation is often not possible with a single marker, and a significant DNA barcoding gap (i.e. interspecific divergence) is often missing between sister taxa in ITS. Further, we can rarely be fully confident about the assigned species-level taxonomy based on available reference sequences. In addition to the inherent limitations, an extra layer of complexity and variation is blended into DNA metabarcoding data due to PCR and sequencing errors that may look similar to natural molecular variation. The bioinformatics processing of ITS amplicons must take into account both the basic properties of the ITS region, as well as the generated errors and biases. In this regard, we cannot adopt approaches and settings from other markers, such as 16S and 18S, blindly. For example, due to intraspecific variability in the ITS region, and sometimes intragenomic variability, ITS sequences must be clustered to approach species level resolution in community studies. Therefore, I argue that the concept of amplicon sequence variants (ASVs) is not applicable. Although the ITS region is by far the best option as a general DNA (meta)barcoding marker for fungi, this contribution is meant to remind against a naive or simplistic use of the ITS region, and for stimulating further discussions.

1. Introduction: the rise of ITS as a marker

The dual revolution of Sanger sequencing (Sanger et al., 1977) and the polymerase chain reaction (PCR) (Mullis and Faloona, 1987; Mullis, 1990), transformed the field of fungal ecology and systematics, starting in the early 1990s (Vilgalys and Hester, 1990; White et al., 1990; Bruns et al., 1991, 1992). Due to its multi-copy nature, ranging from only a few to over 1000 copies in fungal genomes (Lofgren et al., 2019), the ribosomal DNA (rDNA) operon (Fig. 1) was an early target in phylogenetic studies of fungi (Bruns et al., 1990, 1991, 1992), largely because it is easy to amplify, even from tiny amounts of material. Moreover, the rDNA region contains conserved parts suitable for primer design, as well as more variable regions possessing extensive phylogenetic information. Numerous primers still in use today were designed early on to amplify different parts of the rDNA operon (White et al., 1990). While the more slowly evolving SSU and LSU genes are suitable for inferring higher order phylogenies (Berbee and Taylor, 1992; Bruns et al., 1992; Hibbett and Vilgalys, 1993), as well as targets for primer sites, the more rapidly evolving internal transcribed spacer (ITS) is more suitable for genus and species-level analyses (Gardes et al., 1991; Gardes and Bruns, 1993), being generally difficult to align above genus or family-level. Because of its extensive variability, the intergenic spacer (IGS) region was also used in some early population studies (Albee et al., 1996; Guidot et al., 1999). Due to the high activity in fungal evolution and systematics research during the 1990s and 2000s, however, a high number of ITS sequences accumulated in the international nucleotide sequence databases (INSD), such as EMBL or NCBI. Because of greater public database availability, the ITS region emerged as a natural choice as the main fungal DNA barcoding region, formally suggested so in 2012 (Schoch et al., 2012).

In parallel to the implementation of Sanger sequencing in fungal systematics and, somewhat later, in fungal ecology (Horton and Bruns, 2001), there was a rapid development in bioinformatics methods, i.e. on how to process and analyze DNA sequence data. Similarity search algorithms like FASTA (Lipman and Pearson, 1985) and BLAST (Altschul

E-mail address: haavarka@ibv.uio.no.

https://doi.org/10.1016/j.funeco.2023.101274
Received 17 January 2023; Received in revised form 7 May 2023; Accepted 8 June 2023
Available online 26 July 2023

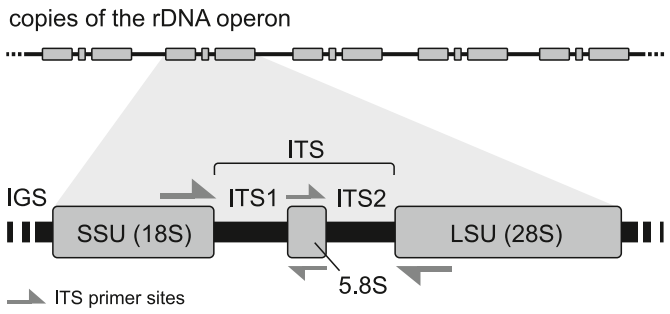

Fig. 1. Sketch of the rDNA operon, with the ITS region indicated. Note that size proportions do not reflect real distances. A high number of primers have been designed for amplification of different parts of the ITS region. A comprehensive overview of ITS primers is provided at the web site of the UNITE database (Abarenkov et al., 2010; Koljalg et al., 2005).

et al., 1990) became parts of the standard toolbox of mycologists, as well as various phylogenetic inferences. FASTA and BLAST enabled researchers to conduct similarity searches against public INSD databases, and in this way perform taxonomic annotation of sequences derived from environmental samples. Although not explicitly named DNA barcoding (or metabarcoding), fungal ecologists performed such analyses from the late 1990s and onwards, although other approaches like PCR-RFLP (Restriction Fragment Length Polymorphism) were more popular to analyze fungal communities early on (Nylund et al., 1995). Environmental sequencing studies from this period relied either on Sanger sequencing of a PCR product amplified directly from environmental DNA (eDNA), e.g. from an ectomycorrhizal root tip, or – if mixed templates were present – cloning followed by Sanger sequencing (O'Brien et al., 2005). Due to the laborious nature of early barcoding techniques and the high expense per sequence, relatively few sequences were typically produced in early environmental sequencing and community ecology studies (Koljalg et al., 2000; Vralstad et al., 2002).

As a discipline, community ecology usually deals with species-level analyses of organisms living at the same place, in order to understand which biotic and abiotic factors and assembly processes structure the community. Hence, DNA markers that approximate species-level resolution are preferable. In this regard, the ITS region represented a more proper choice compared to e.g. the more conserved SSU region, the latter of which provides taxonomic resolution above species or genus level. While the ITS region has been the standard marker for most fungi, parts of the SSU (16S/18S) region have been the standard choice for prokaryotes and non-fungal micro-eukaryotes (Logares et al., 2012). Due to their complex genetics, with multi-nucleate hyphae bearing nuclei with divergent ITS variants, and poorly developed species concepts, the more conserved SSU region have mainly been used for arbuscular mycorrhizal (AM) fungi (Simon et al., 1993; Vandenkoornhuyse et al., 2002).

The number of ITS sequences in INSD databases steadily grew during the 1990s and 2000s, but it was soon realized that many of the reference sequences used for indirect taxonomic annotation of environmental sequences were accessioned under a wrong taxon name and/or with missing or erroneous metadata (Nilsson et al., 2006; Bidartondo et al., 2008). This served as an impetus to develop better-managed reference sequence databases. The UNITE database was established as a response to this need, originally only for ECM fungi (Koljalg et al., 2005), but later widening to the entire fungal kingdom (Abarenkov et al., 2010). By systematic curation and error filtering of INSD ITS sequence accessions, UNITE provides improved reference sequence datasets to the research community (Nilsson et al., 2019). Also other fungal reference databases were developed, tuned towards specific fungal groups and markers, such as MaarjAM for AM fungi (Öpik et al., 2010). The more recently established GlobalFungi database represents a comprehensive collection of fungal ITS high throughput sequence data of global distribution (Vetrovsky et al., 2020).

Around 2008, massively parallelized 'next generation' sequencing (NGS) techniques were introduced to fungal ecology (Buee et al., 2009; Jumpponen and Jones, 2009). During the early NGS phase, 454 sequencing technology dominated, and soon thereafter Illumina sequencing, which provided an even higher throughput albeit somewhat shorter reads (Lindahl et al., 2013). While a single Sanger sequence can cover up to approximately 700 bp and read through the entire ITS region for most fungi, 454 sequences covered up to ~500 bp while Illumina (MiSeq) may cover maximum 350 bp. Due to the limited sequence lengths compared to traditional Sanger sequences, it became necessary to amplify and analyze the ITS1 or the ITS2 regions separately. The new sequencing techniques, producing millions of short sequences in one go, enabled characterization of complex and diverse fungal communities containing mixed DNA templates. With the new sequencing technologies in hand, we were finally at a stage where we could analyze fungal communities in a way that ecologists studying other organisms have done for ages (Peay et al., 2008).

In parallel to the introduction of new DNA sequencing techniques, a suite of new terms were introduced to describe the new high throughput approaches (Taberlet et al., 2012a, 2012b; Coissac et al., 2016). 'DNA barcoding' was suggested for DNA based identification of single physical specimens, whereas 'DNA metabarcoding' refers to the identification of multiple species from a sample with mixed DNA templates. The term 'metagenomics' describes the analyses of all DNA in a sample, while 'metatranscriptomics' deals with analyses of expressed genes through the conversion of RNA to cDNA, followed by sequencing. There is often a misperception among non-practitioners that metagenomics equals metabarcoding or *vice versa*, but these are two very different approaches. This family of logic and inter-related terms (Taberlet et al., 2012a) have been adopted by many researchers, although less precise terms are still used by others for describing DNA metabarcoding, including amplicon sequencing, metagenetics, marker surveys, and the like.

2. Upon closer inspection: inherent limitations in ITS

Ideally, a DNA (meta)barcoding marker should be able to separate between all species in an environmental sample. Accurately doing so involves many factors, but starts with appropriately delineating between-species genetic distances. In other words, there should be a proper *DNA barcoding gap*. The DNA barcoding gap asserts that the variation found across species is higher than the genetic variability within species. Conversely, there should preferably be limited *intraspecific sequence variability*, although this is not a particularly severe problem if a proper barcoding gap is present. In case the marker is a multicopy marker, i.e. repeated within the organisms genomes, there should be no *intragenomic variability* across the copies. Moreover, *the same number of copies* should be present within and across species, making quantitative interpretations more straightforward. Additionally, there should be no *primer mismatches* to the targeted taxonomic groups, no *length variation* in the marker or *variation in GC content* across species, all of this potentially leading to PCR amplification biases. At last, it is preferable if the generated sequences *can be aligned* across the study organisms, to be able to put the organisms and data into a phylogenetic framework. Although being a powerful DNA barcoding marker, the ITS marker does not fulfill any of these criteria. There are numerous inherent limitations associated with the ITS marker to be aware of, as is the case for any other possible DNA marker. Below, I outline the most important limitations of the ITS region as a DNA (meta)barcoding marker.

2.1. Mind the (barcoding) gap

A basic challenge in fungal systematics and DNA metabarcoding, is that it can be difficult to delimit biological species *per se* in fungi. When looking more deeply into the genetic structure of fungal species, many

species turn into species complexes, with sub-populations that possess partially or fully developed intersterility barriers (Taylor et al., 2000, 2006). Species do often not originate through a single speciation event, where intersterility barriers is formed once, but can rather be a stepwise and repeated process towards full intersterility, as has also been the case during human evolution (Bergstrom et al., 2021). Especially in temperate and arctic regions, glacial dynamics have driven species and sub-populations into allopatric distributions in colder periods, followed by sympatric distributions in warmer periods, or sometimes *vice versa* (Taberlet et al., 1998; Hewitt, 1999, 2001). Such global change processes can lead to complex genetic structures within species; introgression between previously geographically separated lineages may lead to non-concordance between gene and species trees (Fig. 2a). Further, in secondary contact (suture) zones, reinforcement can cause accelerated formation of species barriers without high genetic divergence. The ITS marker, like all other parts of the genome, is influenced by population genetic and demographic processes and may have a complex natural history that obscure simple interpretations in DNA metabarcoding studies.

As researchers in fungal taxonomy and systematics have experienced, the interspecific ITS divergence, and hence, the barcoding gap between sister species, varies widely (Fig. 3); in some cases, closely related species share the same ITS sequence (Blaalid et al., 2013; Garnica et al., 2016; Hoang et al., 2019). In such cases, intersterility barriers have evolved more rapidly than the corresponding ITS divergence. This is known to be the case in many genera, such as *Aspergillus* and *Penicillium* (Lucking et al., 2020), *Cortinarius* (Garnica et al., 2016) and *Fusarium* (O'Donnell and Cigelnik, 1997), where additional DNA markers are needed to separate between closely related taxa. In other cases, there can be high sequence divergence between sister-species, and a solid barcoding gap (Nilsson et al., 2008; Lucking et al., 2020). The size of the barcoding gap depends on numerous factors, including the speciation rate in the actual clade, time since speciation and divergence, as well as population level demographic processes. Between closely related and recently diverged species, incomplete lineage sorting may lead to incongruent species and gene trees, where the ITS tree may diverge from the species tree (Fig. 2a). In such cases, multi-locus markers are required to provide better species delimitations. In this regard, the genealogical concordance phylogenetic species recognition (GCPSR) concept can be

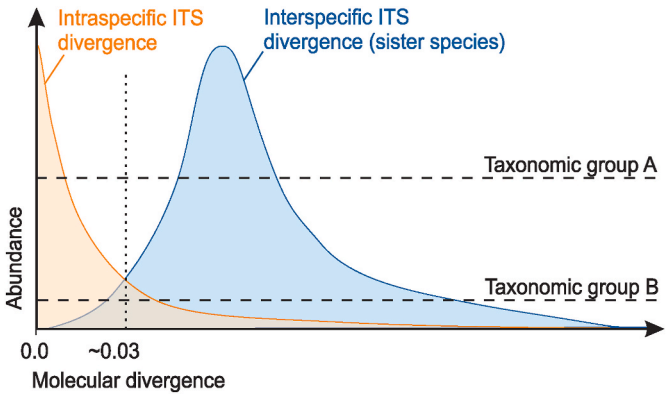

Fig. 3. Conceptual model for the distribution of intraspecific versus interspecific molecular variation in the ITS region, based on information from various studies (Lucking et al., 2020; Nilsson et al., 2008; Schoch et al., 2012; Vu et al., 2019). A much-used cut-off during sequence clustering is 97% sequence similarity; this cut-off might represents the best compromise between lumping and splitting of species (Blaalid et al., 2013). The two taxonomic groups, A and B, indicates two groups were species delimitation would be more (B) and less (A) problematic.

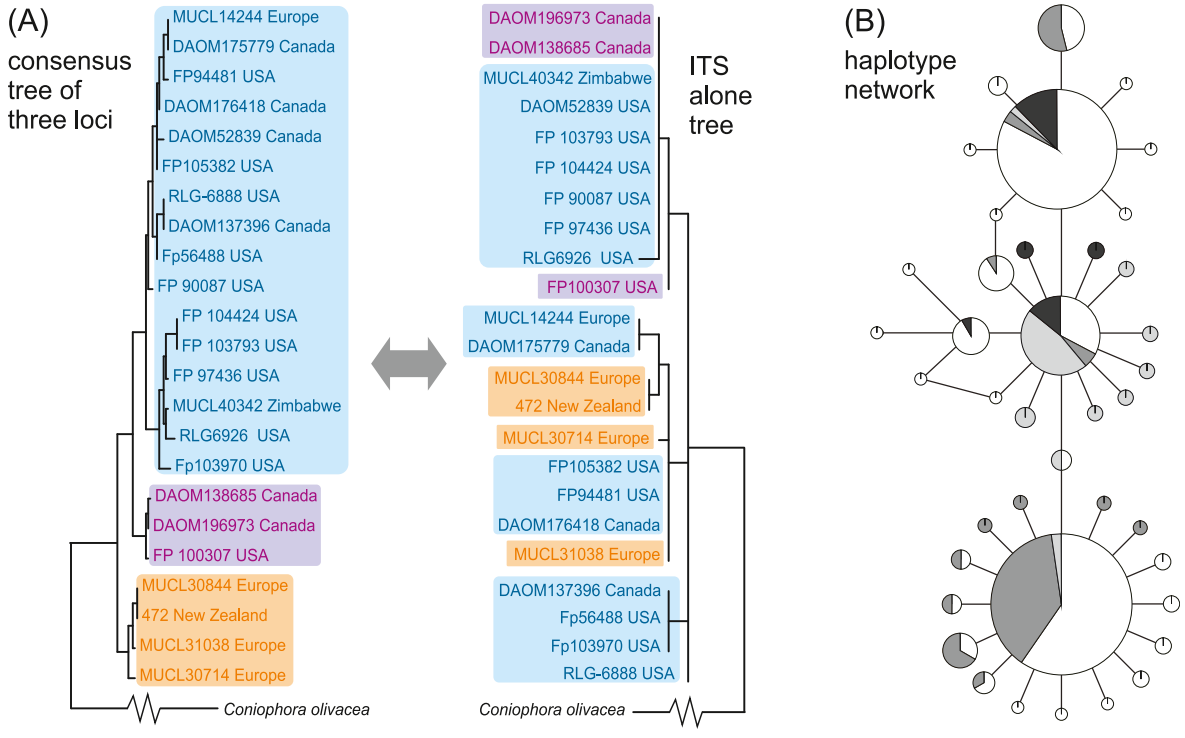

Fig. 2. (A) Schematic trees adopted from an analysis of the *Coniophora arida* species complex (Kauserud et al., 2007a), illustrating the con-concordance between a consensus tree made from three DNA loci (ITS, tef and tub) and the ITS-only tree. In the ITS tree, the three supported groups disintegrates due to incongruence and the lack of an inadequate DNA barcoding gap in the ITS region between the three lineages. (B) A schematic ITS haplotype network of sequences from the wood-decay fungus *Meruliopsis taxicola* drawn based on data from (Skaven Seierstad et al., 2013). Each circle represent an ITS haplotype. The size of the circles represents the haplotype abundance in the overall data and the shadings different continents of origin. As many others, this morphospecies has a complex phylogeographic structure, illustrating that it can be difficult to use ITS to delimit species. The haplotype network also illustrates that ITS sequences must be clustered to approach species-level resolution and that ASVs (i.e. the haplotypes) are not applicable as units in ITS based fungal community analyses.

used to better delimit species (Taylor et al., 2000), although with some limitations (Sukumaran and Knowles, 2017). Unfortunately, multi-locus inferences are not easily achievable based on amplification and analysis of molecular variation from eDNA. It is possible to amplify multiple DNA barcode markers from the same eDNA extract, but it is not straightforward to link data and alleles across unlinked loci from complex eDNA samples, unfortunately, and thus implement a GCPSR approach using eDNA. However, this opportunity could be better evaluated, at least in less complex communities or focusing on some well-known taxonomic groups. In the future, long-read metagenomics or single cell genomics techniques may be able to provide linked multi-locus DNA metabarcoding data, or even better, fully assembled genomes from eDNA, so-called MAGs (Metagenome Assembled Genomes), for more proper species delimitations. The latter approach is used for prokaryotes already (Royo-Llonch et al., 2021), but so far not much for fungi.

2.2. Intraspecific ITS variability

Intraspecific variation in the ITS region (Fig. 2b) are regularly observed in fungal taxa (James et al., 2001; Kauserud et al., 2004, 2007a, 2007b, 2007c; Nilsson et al., 2008; Carlsen et al., 2011; Skaven Seierstad et al., 2013; Estensmo et al., 2021; Seierstad et al., 2021), even at small geographic scales (Kauserud and Schumacher, 2002, 2003a, 2003b; Hughes et al., 2009; Estensmo et al., 2021). Based on these observations, one might speculate that hundreds or thousands of ITS alleles exist in many fungal species. Because of this, the different ITS alleles must be clustered into larger sequence clusters in community studies, to better approximate species level resolution. These clusters are typically known as Operational Taxonomic Units (OTUs) and are used as an approximation for species (or other taxonomic ranks). The OTU term was originally introduced in numerical taxonomy much earlier (Sokal, 1963), as a pragmatic definition of individual specimens grouped by some sort of similarity, phenotypic or genotypic, and was quickly adopted in the field of DNA metabarcoding.

It has been debated what is the best cut-off to use during clustering of sequences into OTUs, to jointly account and fine-tune for intraspecific and interspecific variability. A common threshold adopted from analyses of the SSU region (16S and 18S), is to use 97% sequence identity. However, as important as the threshold itself, is the clustering method. Although not widely implemented, the single-linkage clustering approach seems suitable for accounting for the different levels of intraspecific variability in the ITS region, since larger or smaller sequence clusters, depending on the level of intraspecific sequence variation, can then be constructed. During this type of agglomerative clustering, the clusters are sequentially built up until all sequences linked together by a certain threshold end up in the same cluster. However, the complex phylogeographic structure that exists within many fungal taxa (Fig. 2b) is nevertheless hard to account for, and may lead to over-splitting and overestimation of diversity. Hence, while a missing barcoding gap may lead to clustering of species into the same OTUs (Ryberg, 2015), intraspecific sequence variation may lead to splitting of species into multiple OTUs (Blaalid et al., 2013). In a method study where 26,665 ITS1 and ITS2 sequences of known taxonomy, representing 750 species, were clustered at different thresholds, the *a priori* known number of (morpho)species in the dataset were best re-shaped using 97% sequence similarity threshold during single linkage clustering (Blaalid et al., 2013). Although 97% turned out as a reasonable threshold for both ITS1 and ITS2, many species split into multiple clusters (OTUs) while many sister species were pooled into the same OTU (Blaalid et al., 2013). Obviously, there is no general sequence-clustering threshold across species and there will always be a trade-off between over-splitting and lumping of species. It is possible to correct *post hoc* for over-splitting of taxa in DNA metabarcoding data, though. The program LULU (Froslev et al., 2017), groups genetically similar OTUs into one OTU if they possess the same ecological pattern or distribution across the dataset.

Since numerous ITS alleles often occur within fungal species, it is clearly not suitable to use so-called amplicon sequence variants (ASVs) as units in ITS based fungal community studies. For markers with a high level of intraspecific variation, the diversity will be tremendously overestimated by treating each ITS haplotype as a biological entity in downstream statistical analyses (Fig. 2b). Due to its high sensitivity, it has also been shown that DADA2 splits single bacterial genomes into several ASVs because of intragenomic variability, which is another concern also relevant for the fungal ITS marker (see below). The ASV term is connected to the bioinformatics program DADA2 (Callahan et al., 2016), used to correct for sequencing artifacts in the DNA metabarcoding sequence data. The goal with DADA2 is to identify and separate the original DNA templates from sequencing noise, and remove the latter. Unfortunately, a misconception has been communicated and spread that ASVs are different from OTUs, and should replace OTUs (Callahan et al., 2017). However, ASVs are conceptually nothing else than OTUs; similar to related bioinformatics approaches like Obitools and Swarm (Boyer et al., 2016; Mahe et al., 2014, 2022), DADA2 is ultimately based on agglomeration (clustering) of sequences, where tentative erroneous sequences are assigned to tentative correct, parental sequences. Although branded as an "exact" approach (Callahan et al., 2017), this is not the case. Through the analysis of mock communities, high correspondences have been observed between the original DNA templates and the obtained ASVs (Callahan et al., 2016; Estensmo et al., 2021), but additional ASVs appeared, likely due to PCR or sequencing errors, or alternatively, intragenomic variation. In real samples, with higher richness and complexities compared to mock communities, the DADA2 approach is likely even less exact. The misconception that ASVs and OTUs are conceptually different has led to confusion in the scientific literature, where they often are treated as different units and concepts (Glassman and Martiny, 2018; Tedersoo et al., 2022). However, there is no reason for using a separate term for the units coming from the DADA2 program since the ASV term falls well within the original OTU term. Instead, we should concisely report how the OTUs were generated, whether we use stricter or wider clustering thresholds.

2.3. Intragenomic ITS variability

The rDNA operon is repeated many times in eukaryotic genomes, to enable effective production of ribosomes. In fungi, it was recently shown, based on genome data, that the rDNA copy number ranged between 14 and 1442 across 91 studied fungal taxa (Lofgren et al., 2019). The overall variation among all fungi is likely much higher. In other eukaryotes, like plants and protists, it has been demonstrated that divergent rDNA variants may exist within genomes, either as orthologs or paralogs. Several studies have showed that intragenomic variation in rDNA and ITS also occur in fungal species (Simon and Weiss, 2008; Lindner and Banik, 2011; Lindner et al., 2013; Lucking et al., 2020). Different intragenomic variants may have arisen because of past hybridization events, gene duplications or mutations. Their continued existence within genomes is ultimately due to lack of concerted evolution across the rDNA repeats. Concerted evolution is an evolutionary process that normally homogenizes the rDNA copies within genomes (Elder and Turner, 1995; Ganley and Kobayashi, 2007). The existence of multiple rDNA variants within genomes can in theory obscure DNA metabarcoding analyses and cause inflated species richness estimates. Although widespread in other organisms, it is still somewhat unclear how common or important this phenomenon is among fungi. Much of this variation, if present, likely fall within the range of the normal intraspecific ITS variation and will not dramatically affect the results (Lindner et al., 2013; Lucking et al., 2020). In those cases where more divergent ITS paralogs exist (O'Donnell and Cigelnik, 1997; Aanen et al., 2001; Kauserud and Schumacher, 2003b), this can be adjusted for *post hoc* using e.g. LULU (Froslev et al., 2017) or even by including reference sequences for the different ITS orthologs or paralogs. The increasing number of well-assembled high quality genomes will provide more

information about how widespread intragenomic ITS variation is among fungi, although re-analysis of the raw reads might be needed (Lofgren et al., 2019).

Noteworthy, intragenomic variation should not be confounded with intra-individual variation, which may exist in e.g. a heterokaryotic basidiomycete or an AM fungus, or a diploid chytrid fungus. In such fungi, ITS polymorpshisms and heterozygous sites can often be observed in a single mycelium or genet, since more than one genome and hence, allelic variation may be present.

As outlined above, the overall molecular variation you observed in the ITS marker in a community study is organized and nested at different levels, ranging from (1) *interspecific* variation, (2) *intraspecific* variation, (3) *intra-individual* variation between different nuclei/genomes, and (4) *intragenomic* variation.

2.4. PCR biases: copy number variation, AT/GC content, length variation and primer mismatch

Several inherent ITS characteristics may cause biases during PCR amplification. Linked to the discussion above, species and genomes with a high copy number will be more rapidly amplified during PCR and hence, become proportionally more abundant in the final dataset compared to species with a low copy number. The same is probably the case for ITS sequences with a high AT content, as compared to GC content; it might be expected that double-stranded DNA denatures more easily during PCR if the AT content is relatively high, due to a lower melting temperature. The AT/GC content varies extensively across fungal species and lineages (Yang et al., 2018), and may contribute to amplification bias. Analogous to this, shorter ITS sequences will be easier amplified than longer sequences. In ITS2, there is a systematic length biases across fungal phyla (Bellemain et al., 2010; Yang et al., 2018), where Ascomycota in general possess significantly shorter sequences than Basidiomycota. Hence, in metabarcoding studies using ITS2 as a marker, ascomycetes will be preferentially amplified. The same length bias will be present during amplification of the entire ITS region, which is now achievable with third generation sequencing techniques. Length bias does not only come into play during PCR; a significant length bias may also be introduced during the sequencing process. During Illumina sequencing, shorter fragments are more frequently sequenced than longer fragments (Castano et al., 2020). Hence, in DNA metabarcoding studies of the ITS2 region using the Illumina platform, a double length bias is likely introduced, both during PCR and sequencing. Consequently, the proportion of taxa and lineages with shorter ITS2 sequences may look far more abundant in the final data than what they actually are in the original DNA extracts. Also third generation sequencing techniques like PacBio have length bias, going in the opposite direction, though. Here, longer sequences may load better than shorter during sequencing, at least under a certain sequence length threshold. Another problem when analyzing the entire ITS region or ITS1, is the presence of introns towards the end of SSU, where many of the ITS primers are located (Bhattacharya et al., 2000; Tedersoo and Lindahl, 2016). This is especially prevalent in some lineages of Ascomycota (Bhattacharya et al., 2000), and may lead to instances where important community members not are amplified and sequenced. Due to the considerable length variation in the ITS region, the bioinformatics analyses must allow for different lengths and not e.g. truncate them to uniform lengths.

Severe amplification biases may also be introduced due to lineage-specific mutations in the primer regions (Schadt and Rosling, 2015; Tedersoo and Lindahl, 2016). Most of the commonly used ITS primers are known to have biases towards certain taxonomic groups (Bellemain et al., 2010). For example, the widely distributed Archaeorhizomycetes was often missed out in ITS surveys, due to the presence of a mutation (2 bp inversion) in the commonly used ITS4 primer site (Schadt and Rosling, 2015). To better avoid primer biases, degenerate primers, i.e. a mixture of different primers, can be used, although this may complicate the PCR setup. To better understand the effect of the biases, mock communities with ITS sequences of different lengths and GC-content should be included and analyzed (Castano et al., 2020). Additional *in silico* studies also have a role to play here, for example by examining the specificity and coverage of primers using sequence databases, and their physical properties (e.g. Bellemain et al., 2010).

3. Going beyond ITS: general technical errors and biases

High throughput sequencing have provided novel opportunities to scrutinize fungal communities. However, with the new methods came also new challenges, errors and biases beyond the inherent limitations discussed above. Technical errors and biases we must deal with include *PCR and sequencing mutations*, *chimeric sequences*, *tag-jumping* and other *contamination* problems, as well as differences in *sequencing depth* across samples. These issues are of more general nature and not connected to ITS as a DNA barcode marker, *per se*. However, since some of these technical errors, such as PCR and sequencing mutations, imitate true biological patterns and ITS variation, it is important to consider them together. Given all putative errors and biases, another challenging is to what degree we can use metabarcoding data quantitatively.

3.1. PCR and sequencing mutations

Sequences derived from second generation sequencing techniques, such as Illumina, include in general far more errors compared to traditional Sanger sequences, at least when the latter is obtain through direct Sanger sequencing (and not through cloning). During PCR, a small proportion of the nucleotides in the template DNA is erroneously translated by the polymerase enzyme, even when using proofreading enzymes, leading to PCR-generated mutations (Chen et al., 1991; Potapov and Ong, 2017). Certain polymerases also have a primer editing capacity, which can be used to rescue drop-out of taxa with primer mismatches (Gohl et al., 2021). PCR mutations happen rarely; error rates of Taq polymerase ranges between 0.1% and 0.001% (Chen et al., 1991; Potapov and Ong, 2017), but will nevertheless have a great impact when millions of sequences are generated and sequenced. Even a small error rate will lead to many spurious mutations and sequences deviating from the original one with one or a few base pairs. Unfortunately, molecular variation generated by PCR-errors look similar to true molecular variation, with abundant haplotypes representing the true (parental) haplotypes surrounded by multiple rare ones in star shaped patterns (Fig. 2b). Although impossible to avoid completely, there are ways to minimize PCR errors, including the usage of proofreading enzymes, and possibly lowering the number of cycles; see Tedersoo et al. (2022) for a broader discussion.

During direct Sanger sequencing, PCR mutations generally do not become visible, since the resulting chromatogram is made up of an average signal of a high number of DNA fragments/amplicons. The original correct sequence(s) tends to dominate quantitatively, since the original template(s) attends most PCR cycles. In a Sanger chromatogram, the PCR-errors may just appear as noise in the bottom part of the chromatogram. On the contrary, during next generation sequencing, as well as traditional clone-based Sanger sequencing, the final sequences originates from single fragments that might include error(s) introduced during PCR. Hence, since a single DNA molecule gives rise to the resulting sequence, the PCR errors become visible and are not masked by the correct majority. These errors must be accounted for in the later bioinformatics workflow. Similar types of errors are also introduced during the sequencing step, somewhat dependent on the sequencing technique. While Illumina sequencing generates an additional layer of errors, circular consensus sequencing on the PacBio platform (with a high number of passes) provides high quality sequences with few additional errors (Castano et al., 2020). Hence, while the sequencing error rate varies greatly with different sequencing methods, PCR errors will be visible even when producing what appear to be high quality

sequences. Early DNA metabarcoding studies relied on inferior bioinformatics approaches not able to correct properly for these errors, likely resulting in inflated levels of OTUs. Luckily, new bioinformatics approaches have been developed to better tackle PCR and sequencing errors (Mahe et al., 2014; Boyer et al., 2016; Callahan et al., 2016). However, it is still very demanding to separate *de facto* molecular ITS variation from PCR and sequencing errors, especially since they to a large extent look the same (see Fig. 2b).

The original way to account for PCR and sequencing errors, as well as intraspecific variability in ITS, in one go, was to cluster the sequences into OTUs based on a certain level of sequence similarity, say 97%. This is a crude approach where the different sources of molecular ITS variation not is considered separately. Luckily, more advanced and precise ways for accounting for PCR and sequencing errors have been developed, including the algorithms and approaches implemented in the software packages Obitools (Boyer et al., 2016), Swarm (Mahe et al., 2022) and DADA(2) (Callahan et al., 2016). These programs build, to some extent, on the same philosophy accounting for errors; they aim to recognize and differentiate between the putative correct, parental sequences reflecting the original DNA templates, and those where artificial mutations have been introduced during PCR and/or sequencing. The putative artificially mutated sequences are identified and assigned to the putative correct sequences based on certain principles and algorithms differing among the programs. In general, the correct sequences are expected to be more abundant in the final sequence pool compared to the artificial sequences, since the original correct sequences attend more PCR cycles. Further, the parental sequences are expected to be very similar to the mutated sequence(s), most of them differing by only one bp. These programs are able to prune away the most likely PCR mutations (Callahan et al., 2019; Estensmo et al., 2021). When working perfectly, the user will end up with a set of sequences perfectly matching the original haplotypes or alleles in the PCR mix. With the DADA(2) program (Callahan et al., 2017, 2019), the term Amplicon Sequence Variant (ASV) was introduced, which is basically a novel term for the inferred, putative correct, original sequence. However, as pinpointed above, DADA(2) also involves agglomeration (clustering) of sequences, and do not represents anything else than some of the other approaches, conceptually.

3.2. Chimeras

Unfortunately, we also have to deal with other types of technical errors; when mixed templates are present in the sample and the PCR reaction, these may combine into artificial chimeric sequences (chimeras) (Kopczynski et al., 1994; Wang and Wang, 1996). Again, these errors do not become visible with traditional direct Sanger sequencing. A chimeric sequence is formed when a DNA fragment from an aborted extension during the PCR anneals to a single-stranded DNA template of different origin in a downstream PCR cycle and acts as a primer (Kopczynski et al., 1994). In this way an artificial sequence derived from two (or more) templates is formed. Somewhat counterintuitive, proofreading enzymes have a tendency to generate more chimeric sequences than non-proofreading enzymes (Ahn et al., 2012). Chimeric sequences will stand out as divergent sequences, often placed basally or on long branches in phylogenetic trees, and will inflate the species richness if not corrected for (Engelbrektson et al., 2010; Kunin et al., 2010).

In smaller datasets and sequence alignments, chimeras can be manually spotted and removed. In smaller datasets, it is also possible to manually construct phylogenetic trees from different parts of the alignment and then look for incongruent OTUs that shift position across threes, indicating recombination. However, in extensive metabarcoding datasets it is impractical to do such manual inspections, especially when dealing with the highly variable ITS region. Therefore, different programs and algorithms have been developed to automatically detect and flag putative chimeric sequences, including chimeraslayer (Haas et al., 2011), uchime (Edgar et al., 2011), and perseus (Quince et al., 2011). These programs use either a reference database containing an alignment of chimera-free sequences (Haas et al., 2011), or a *de novo* approach, using the input data for detecting chimeras (Edgar et al., 2011; Quince et al., 2011). A comprehensive ITS sequence dataset for reference-based chimera checking is available (Nilsson et al., 2015). However, reference based analyses only works well when you have a well-propagated reference database, which do not exist for fungi, where only a small proportion of the real diversity is represented in the reference databases (Lucking et al., 2020). Hence, for fungi, *de novo* chimera checking is preferable, where sequences in the same dataset are compared against each other (Aas et al., 2017). One much used strategy is to break the sequences into smaller pieces and conduct independent similarity searches with the different parts; if the different parts of a sequence matches better to different divergent sequences in the dataset, rather to itself, the sequence is flagged as a putative chimeric sequence. Some limitations are built into most of the chimera checking programs, including the assumption that there is only two parent sequences. A common assumption is also that chimeric sequence are rarer than its parental sequences (Edgar et al., 2011), since they have attended fewer PCR cycles.

The chimera formation rate increases with level of template similarity (Shin et al., 2014; Aas et al., 2017), since the polymerase enzyme more easily combines similar templates. This means that far less chimeric sequences are formed when targeting the variable ITS region compared to e.g. the SSU or LSU regions (Jumpponen and Johnson, 2005; Aas et al., 2017). When using automated bioinformatics pipelines like uchime (Edgar et al., 2011) to detect chimeras, a severe problem may be the high false positive rate of chimera detection. In a method study using mock communities, it was observed that most of the automatically flagged ITS chimeras were false positives (Aas et al., 2017). False positives might be especially prevalent in ITS datasets, since some level of natural intra-locus recombination likely is prevalent (Kauserud and Schumacher, 2003b). Such recombination events can easily be flagged as chimeras. Changing the parameter settings in the chimera detection programs are probably needed to better fine-tune them towards the variable ITS marker; the standard settings based on 16S or 18S data are clearly not suitable and may lead to the removal of real diversity (Aas et al., 2017). A very stringent approach to avoid chimeras is to remove OTUs only appearing in one sample, with the assumptions that the same chimera is not formed twice. Although more labor-intensive and expensive, a better strategy is to independently tag and run duplicate or triplicate PCRs of each sample, and only accept sequences generated independently in the PCR replicates for further analyses.

3.3. Tag jumping and contaminations

During DNA metabarcoding analyses, a high number of samples and amplicons are typically pooled into one or several libraries and sequenced together. To be able to keep the samples apart, each amplicon must be tagged with a short oligo. In a method study, where ITS amplicons were generated and sequenced from numerous single spore cultures (Carlsen et al., 2012), only ITS sequences representing the single species were expected in each sample. However, a small amount of sequences from other species had leaked across samples; the samples were cross contaminated (Carlsen et al., 2012). This was caused by tag jumping, also mentioned as tag leakage or bleeding. This problem, tentatively causing a high number of false positives, was soon detected in other studies (Esling et al., 2015; Schnell et al., 2015). In extreme cases, up to 28% of the sequences had shifted tags due to tag jumping (Esling et al., 2015). Most of the tag jumping seemingly happened during the final library preparation step, where a few PCR cycles are used to link on sequencing adapters. Hence, avoiding this final PCR step, instead using ligation, seems to reduce the tag-jumping problem considerably (Taberlet et al., 2018). Furthermore, by tagging each amplicon in both ends with a unique tag combination, it is possible to detect and remove sequences undergoing tag jumping from the final

dataset (Carlsen et al., 2012). DNA metabarcoding analyses are highly vulnerable to other types of contaminations that can happen during sampling, DNA extraction, PCR or sequencing. The inclusion of multiple negative and positive controls, including biological or synthetic mock communities (Palmer et al., 2018), is highly advisable to be able to handle contamination problems in a best possible manner (Zinger et al., 2019).

3.4. Sequencing depth

In DNA metabarcoding studies, we typically aim to obtain a similar number of sequences per sample, independent of the amount of DNA in the starting material or DNA extracts. To achieve this, we need to normalize the amount of ITS templates across samples and preferably obtain equimolar ratios in the final sequencing library. However, this is a challenging task whether you normalize based on the PCR bands' gel intensities, spectrophotometer analyses or a normalization kit. The amount of sequences across samples often varies 10-fold, which can partly also be attributed to factors discussed above, such as length bias introduced during sequencing. Calculating sample based OTU accumulation curves is a good first step to assess to what degree the diversity has been covered. While beta diversity patterns can be highly stable and robust with different sequencing depths and data treatments (Botnen et al., 2018), alpha and gamma diversity may be heavily influenced by sequencing depth in case the OTU accumulation curves have not plateaued. There is an ongoing debate about what are the most suitable options to correct for sequencing depth differences (McMurdie and Holmes, 2014; Weiss et al., 2017; McKnight et al., 2018). Simplistic approaches, like turning your data into proportional data or rarifying down to a common number of sequences per sample have been criticized, and statistically more advanced approaches suggested (McMurdie and Holmes, 2014). However, for community-level comparisons, others strongly advocate for simple rarefication or the use of proportional data (McKnight et al., 2018). As often is the case, the best solution might be data and context-dependent (Weiss et al., 2017).

3.5. Quantification

Obtaining better quantitative data can be seen as the *holy grail* in DNA metabarcoding. Unfortunately, most of the current DNA metabarcoding methodologies do not provide reliable quantitative data (Luo et al., 2022; Shelton et al., 2022). The obtained sequence counts of the different species are likely poor indicators of absolute quantities in the starting material due to the many biases that come into play during sampling, DNA extraction, PCR and sequencing, as well as the different biological and genetic properties of the species. Quantitative comparisons between species, both within and especially between samples, are extremely challenging due to e.g. varying extraction efficiency, varying number of nuclei and rDNA operons, length and GC content differences causing PCR biases - just to name a few out of many influencing factors. Quantitative comparisons of the same species across samples is also challenging, but more achievable. One suggested approach is to introduce a DNA spike-in to every sample before DNA extraction or just after, and use the number of sequences from the spike-in to calibrate across samples (Stammler et al., 2016; Luo et al., 2022). Another possibility is to conduct qPCR in parallel, for calibration (Shelton et al., 2022). An interesting innovation is to tag each original template with unique molecular identifiers (UMIs), in theory enabling tracking all the generated sequences back to the original template (Fields et al., 2020). However, the implementation of UMIs introduces extra steps during library preparation that likely are associated with extra biases or errors.

4. Taxonomic annotation and ITS reference databases

Among the final steps in the DNA metabarcoding workflow is taxonomic annotation of the OTUs using e.g. UNITE or INSD databases. In UNITE, the ITS reference sequences are organized into species hypotheses through sequence clustering (Koljalg et al., 2013). Taxonomic annotation can be done in different ways, but all involve the use of reference sequences with known taxonomy. The simplest approach is to use BLAST (Altschul et al., 1990) and evaluate the top hit(s) to assign taxonomy. In this regard, the level of identity and coverage must be taken into consideration and reported. However, even when having 100% identity and coverage with a certain reference sequence, species hypothesis or taxon, we cannot ultimately be fully confident that our environmental sequence represents this taxon. For example, a known or unknown sister taxon with a similar ITS sequence might not be accessioned in the reference database, or incomplete lineage sorting can break the gene-tree species-tree congruency (Fig. 2a) and obscure taxonomic assignment.

There are, however, more elaborated ways to assign taxonomy, also providing estimates of the uncertainty involved. Programs like the RDP classifier (Wang et al., 2007) and Sintax (Edgar, 2016), offers heuristic estimates of the reliability of the taxonomic assignment. The program Protax (Somervuo et al., 2016), including the fungal-specific version Protax-fungi (Abarenkov et al., 2018), performs probabilistic identification and apparently performs better than RDP and Sintax for datasets poorly covered by the reference databases (Abarenkov et al., 2018). Protax-fungi also returns classification probabilities at multiple taxonomic ranks. However, assigning environmental sequences to different taxonomic ranks, a purely human construct, is a challenging task with poorly developed reference databases. Another alternative is to map environmental sequences by so-called phylogenetic binning, using multiple alignments and backbone phylogenies (Berger et al., 2011) or a closed reference approach (Cline et al., 2017). Since reliable ITS alignments and phylogenies hardly can be obtained above the genus level, phylogenetic binning does not seems applicable in most cases, unless focusing on a limited taxonomic group. However, long-read sequencing techniques, enabling the inclusion of LSU and SSU data in the same sequence, provide better opportunities for reliable phylogenies and phylogenetic binning. Moreover, the closed reference approach, where the non-clustered sequences are directly mapped to the reference sequences (Cline et al., 2017), demands a well-propagated reference database far beyond what is currently available. Still, less than one percent of the estimated global fungal diversity of ~6.3 million species (Baldrian et al., 2022) are represented in reference sequence databases, or 1–2% when considering 2.2–3.8 million species (Lucking et al., 2020). However, with time, this less computational intensive approach will likely become even more relevant.

5. Future perspectives on the use of ITS

Despite its limitations, no viable alternatives to ITS have emerged so far for broad community analyses of fungi. Alternative DNA markers shared across the fungal tree of life, situated in central housekeeping genes, have been searched for (Stielow et al., 2015; Vetrovsky et al., 2016), but many candidates are hampered by mutations in the codons third bp positions, making primer design difficult across broader taxonomic groups. However, for some taxonomic groups, markers situated in e.g. the *tef* and *rpb* genes are viable options (Stielow et al., 2015; Vetrovsky et al., 2016). Even though the ITS region seems to be the best choice, it possess numerous limitations that should not be ignored, including lack of a barcoding gap, and the varying levels of intraspecific and interspecific molecular variation. PCR errors are also blended in during the lab-work, and we should therefore be very careful about interpreting DNA metabarcoding data as exact information (Callahan et al., 2016). However, some of the weaknesses of the ITS marker represent at the same times its strengths, indicating a trade-off. For example, its multi-copy nature makes it easier to amplify but enables at the same time intragenomic variation and also complicates quantitative assessments since the copy-number varies widely.

How much does the inherent limitations and the technical errors

matters, and when? The short answer is - it depends. The inherent limitations and technical errors may have fundamentally different effects based on which responses you are analyzing. Several studies have shown that beta diversity patterns, i.e. how the overall fungal communities changes along environmental gradients or with different treatments, are highly robust to different data treatments and, likely, PCR and sequencing errors (Botnen et al., 2018; Glassman and Martiny, 2018; Lekberg et al., 2014). A method study demonstrated that fungal communities mainly are structured by a few dominant OTUs, and that these OTUs are stable and robust towards different data treatment (Botnen et al., 2018). In this study, multiple fungal ITS datasets from different environments were clustered across a wide range of similarity cut-offs (87–99%). The observed community structure largely remained the same across data treatments of the above-mentioned reason. Removing most of the OTUs from the datasets, except the most abundant ones, did not influence the observed community structure (Botnen et al., 2018). In light of this, one might argue that some of the highlighted limitations of the ITS marker, including lack of barcoding gap and high intraspecific variation, often not are relevant in a real-world setting. For example, sister taxa are often not present at the same place or in the same study, hence, the missing barcoding gap becomes less relevant. In addition, the intraspecific ITS variation is expectedly smaller in spatially local-scale studies compared to broad-scale studies. Furthermore, functional ecology is usually inferred at the genus level, so exact species identity may be of less importance. On the contrary, when looking at absolute richness or diversity patterns as such, the results will be highly sensitive to data treatment, since the number of clusters and OTUs will vary widely depending on data treatment. Moreover, due to primer mismatches and length issues, many fungal groups are often not amplified during PCR, which must be taken into account.

For how long will the ITS region survive as a general single-locus marker, before long-read sequencing techniques and other metagenomics approaches makes it partly obsolete? Although long-read sequencing techniques now provide the opportunity to generate very long and high quality sequences (Tedersoo et al., 2021), covering e.g. the entire rDNA operon (including ITS) (Wurzbacher et al., 2019), it will still remain more challenging to PCR amplify a longer DNA region from most environmental samples compared to a shorter, limiting the ability to detect all the fungi in a sample, and especially so if the DNA is partly degraded. Different PCR biases may also have a stronger influence on longer markers. When it comes to metagenomics, i.e. the sequencing of all DNA in an environmental sample, it is still a challenging task to cover and assemble the DNA in complex environmental samples, like soil, and at the same time keep a well-replicated study design. Metagenomics is of limited use when the target organisms are relatively rare; in these cases, enormous sequencing depth would be required. Technological progress in single-cell genomics, transcriptomics and microfluidics techniques will impact fungal ecology research and new approaches may replace some of the current uses of the ITS marker. However, for many purposes, like a rapid and more low-tech taxonomic screening of fungal communities, or for the analyses of partly degraded or ancient DNA, ITS will likely be a relevant marker for a long time. The ITS marker can also serve as a tool to select the most informative samples for more detailed and expensive analyses, including attempts to assemble MAGs or conduct in-depth transcriptome analyses. Ultimately - and independent of the methodological approach itself - we need more extensive reference sequence databases, whether they are only covering ITS, the full rDNA operon or entire genomes. For better propagating the databases, fungal ecologists rely on the continuous work in fungal taxonomy and systematics.

Acknowledgements

Prof. Peter Kennedy, Prof. Ari Jumpponen and two anonymous reviewers are acknowledged for valuable input to the manuscript, and different personnel in OMG (Oslo Mycology Group) for discussions on these topics throughout the years.

References

Aanen, D.K., Kuyper, T.W., Hoekstra, R.F., 2001. A widely distributed ITS polymorphism within a biological species of the ectomycorrhizal fungus Hebeloma velutipes. Mycol. Res. 105, 284–290.

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. Mol Ecol Resour 17, 730–741.

Abarenkov, K., Nilsson, R.H., Larsson, K.H., Alexander, I.J., Eberhardt, U., Erland, S., Hoiland, K., Kjoller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A.F.S., Tedersoo, L., Ursing, B.M., Vralstad, T., Liimatainen, K., Peintner, U., Koljalg, U., 2010. The UNITE database for molecular identification of fungi - recent updates and future perspectives. New Phytol. 186, 281–285.

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

Ahn, J.H., Kim, B.Y., Song, J., Weon, H.Y., 2012. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. J. Microbiol. 50, 1071–1074.

Albee, S.R., Mueller, G.M., Kropp, B.R., 1996. Polymorphisms in the large intergenic spacer of the nuclear ribosomal repeat identify Laccaria proxima strains. Mycologia 88, 970–976.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Baldrian, P., Vetrovsky, T., Lepinay, C., Kohout, P., 2022. High-throughput sequencing view on the magnitude of global fungal diversity. Fungal Divers. 114, 539–547.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., Kauserud, H., 2010. ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiol. 10.

Berbee, M.L., Taylor, J.W., 1992. Detecting morphological convergence in true fungi, using 18S rRNA gene sequence data. Biosystems 28, 117–125.

Berger, S.A., Krompass, D., Stamatakis, A., 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Syst. Biol. 60, 291–302.

Bergstrom, A., Stringer, C., Hajdinjak, M., Scerri, E.M.L., Skoglund, P., 2021. Origins of modern human ancestry. Nature 590, 229–237.

Bhattacharya, D., Lutzoni, F., Reeb, V., Simon, D., Nason, J., Fernandez, F., 2000. Widespread occurrence of spliceosomal introns in the rDNA genes of ascomycetes. Mol. Biol. Evol. 17, 1971–1984.

Bidartondo, M.I., Bruns, T.D., Blackwell, M., Edwards, I., Taylor, A.F.S., Horton, T., Zhang, N., Koljalg, U., May, G., Kuyper, T.W., Bever, J.D., Gilbert, G., Taylor, J.W., DeSantis, T.Z., Pringle, A., Borneman, J., Thorn, G., Berbee, M., Mueller, G.M., Andersen, G.L., Vellinga, E.C., Branco, S., Anderson, I., Dickie, I.A., Avis, P., Timonen, S., Kjoller, R., Lodge, D.J., Bateman, R.M., Purvis, A., Crous, P.W., Hawkes, C., Barraclough, T., Burt, A., Nilsson, R.H., Larsson, K.H., Alexander, I., Moncalvo, J.M., Berube, J., Spatafora, J., Lumbsch, H.T., Blair, J.E., Suh, S.O., Pfister, D.H., Binder, M., Boehm, E.W., Kohn, L., Mata, J.L., Dyer, P., Sung, G.H., Dentinger, B., Simmons, E.G., Baird, R.E., Volk, T.J., Perry, B.A., Kerrigan, R.W., Campbell, J., Rajesh, J., Reynolds, D.R., Geiser, D., Humber, R.A., Hausmann, N., Szaro, T., Stajich, J., Gathman, A., Peay, K.G., Henkel, T., Robinson, C.H., Pukkila, P. J., Nguyen, N.H., Villalta, C., Kennedy, P., Bergemann, S., Aime, M.C., Kauff, F., Porras-Alfaro, A., Gueidan, C., Beck, A., Andersen, B., Marek, S., Crouch, J.A., Kerrigan, J., Ristaino, J.B., Hodge, K.T., Kuldau, G., Samuels, G.J., Raja, H.A., Voglmayr, H., Gardes, M., Janos, D.P., Rogers, J.D., Cannon, P., Woolfolk, S.W., Kistler, H.C., Castellano, M.A., Maldonado-Ramirez, S.L., Kirk, P.M., Farrar, J.J., Osmundson, T., Currah, R.S., Vujanovic, V., Chen, W.D., Korf, R.P., Atallah, Z.K., Harrison, K.J., Guarro, J., Bates, S.T., Bonello, P., Bridge, P., Schell, W., Rossi, W., Stenlid, J., Frisvad, J.C., Miller, R.M., Baker, S.E., Hallen, H.E., Janso, J.E., Wilson, A.W., Conway, K.E., Egerton-Warburton, L., Wang, Z., Eastburn, D., Ho, W. W.H., Kroken, S., Stadler, M., Turgeon, G., Lichtwardt, R.W., Stewart, E.L., Wedin, M., Li, D.W., Uchida, J.Y., Jumpponen, A., Deckert, R.J., Beker, H.J., Rogers, S.O., Xu, J.A.P., Johnston, P., Shoemaker, R.A., Liu, M.A., Marques, G., Summerell, B., Sokolski, S., Thrane, U., Widden, P., Bruhn, J.N., Bianchinotti, V., Tuthill, D., Baroni, T.J., Barron, G., Hosaka, K., Jewell, K., Piepenbring, M., Sullivan, R., Griffith, G.W., Bradley, S.G., Aoki, T., Yoder, W.T., Ju, Y.M., Berch, S. M., Trappe, M., Duan, W.J., Bonito, G., Taber, R.A., Coelho, G., Bills, G., Ganley, A., Agerer, R., Nagy, L., Roy, B.A., Laessoe, T., Hallenberg, N., Tichy, H.V., Stalpers, J., Langer, E., Scholler, M., Krueger, D., Pacioni, G., Poder, R., Pennanen, T., Capelari, M., Nakasone, K., Tewari, J.P., Miller, A.N., Decock, C., Huhndorf, S., Wach, M., Vishniac, H.S., Yohalem, D.S., Smith, M.E., Glenn, A.E., Spiering, M., Lindner, D.L., Schoch, C., Redhead, S.A., Ivors, K., Jeffers, S.N., Geml, J., Okafor, F., Spiegel, F.W., Dewsbury, D., Carroll, J., Porter, T.M., Pashley, C., Carpenter, S.E., Abad, G., Voigt, K., Arenz, B., Methven, A.S., Schechter, S., Vance, P., Mahoney, D., Kang, S.C., Rheeder, J.P., Mehl, J., Greif, M., Ngala, G.N., Ammirati, J., Kawasaki, M., Gwo-Fang, Y.A., Matsumoto, T., Smith, D., Koenig, G., Luoma, D., May, T., Leonardi, M., Sigler, L., Taylor, D.L., Gibson, C., Sharpton, T., Hawksworth, D.L., Dianese, J.C., Trudell, S.A., Paulus, B., Padamsee, M., Callac, P., Lima, N., White, M., Barreau, C., Juncai, M.A., Buyck, B., Rabeler, R.K., Liles, M.R., Estes, D., Carter, R., Herr, J.M., Chandler, G., Kerekes, J., Cruse-Sanders, J., Marquez, R.G., Horak, E., Fitzsimons, M., Doring, H., Yao, S., Hynson, N., Ryberg, M., Arnold, A.E., Hughes, K., 2008. Preserving accuracy in GenBank. Science 319, 1616, 1616.

Blaalid, R., Kumar, S., Nilsson, R.H., Abarenkov, K., Kirk, P.M., Kauserud, H., 2013. ITS1 versus ITS2 as DNA metabarcodes for fungi. Mol Ecol Resour 13, 218–224.
Botnen, S.S., Davey, M.L., Halvorsen, R., Kauserud, H., 2018. Sequence clustering threshold has little effect on the recovery of microbial community structure. Mol Ecol Resour 18, 1064–1076.
Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. OBITOOLS: a UNIX-inspired software package for DNA metabarcoding. Mol Ecol Resour 16, 176–182.
Bruns, T.D., Fogel, R., Taylor, J.W., 1990. Amplification and sequencing of DNA from fungal herbarium specimens. Mycologia 82, 175–184.
Bruns, T.D., White, T.J., Taylor, J.W., 1991. Fungal molecular systematics. Annu. Rev. Ecol. Systemat. 22, 525–564.
Bruns, T.D., Vilgalys, R., Barns, S.M., Gonzalez, D., Hibbett, D.S., Lane, D.J., Simon, L., Stickel, S., Szaro, T.M., Weisburg, W.G., Sogin, M.L., 1992. Evolutionary relationships within the fungi: analyses of nuclear small subunit rRNA sequences. Mol. Phylogenet. Evol. 1, 231–241.
Buee, M., Reich, M., Murat, C., Morin, E., Nilsson, R.H., Uroz, S., Martin, F., 2009. 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. New Phytol. 184, 449–456.
Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581.
Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11, 2639–2643.
Callahan, B.J., Wong, J., Heiner, C., Oh, S., Theriot, C.M., Gulati, A.S., McGill, S.K., Dougherty, M.K., 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Res. 47, e103.
Carlsen, T., Engh, I.B., Decock, C., Rajchenberg, M., Kauserud, H., 2011. Multiple cryptic species with divergent substrate affinities in the Serpula himantioides species complex. Fungal Biol-Uk 115, 54–61.
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 Ecol 5, 747–749.
Castano, C., Berlin, A., Durling, M.B., Ihrmark, K., Lindahl, B.D., Stenlid, J., Clemmensen, K.E., Olson, A., 2020. Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities. New Phytol. 228, 1149–1158.
Chen, J., Sahota, A., Stambrook, P.J., Tischfield, J.A., 1991. Polymerase chain-reaction amplification and sequence-analysis of human mutant adenine phosphoribosyltransferase genes - the nature and frequency of errors caused by Taq DNA-polymerase. Mutat. Res. 249, 169–176.
Cline, L.C., Song, Z.W., Al-Ghalith, G.A., Knights, D., Kennedy, P.G., 2017. Moving beyond de novo clustering in fungal community ecology. New Phytol. 216, 629–634.
Coissac, E., Hollingsworth, P.M., Lavergne, S., Taberlet, P., 2016. From barcodes to genomes: extending the concept of DNA barcoding. Mol. Ecol. 25, 1423–1428.
Edgar, R.C., 2016. SINTAX: a Simple Non-bayesian Taxonomy Classifier for 16S and ITS Sequences. BioRxiv.
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.
Elder, J.F., Turner, B.J., 1995. Concerted evolution of repetitive DNA-sequences in eukaryotes. Q. Rev. Biol. 70, 297–320.
Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., Hugenholtz, P., 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J. 4, 642–647.
Esling, P., Lejzerowicz, F., Pawlowski, J., 2015. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. Nucleic Acids Res. 43, 2513–2524.
Estensmo, E.L.F., Maurice, S., Morgado, L., Martin-Sanchez, P.M., Skrede, I., Kauserud, H., 2021. The influence of intraspecific sequence variation during DNA metabarcoding: a case study of eleven fungal species. Mol Ecol Resour 21, 1141–1148.
Fields, B., Moeskjær, S., Friman, V.-P., Andersen, S.U., Young, J.P.W., 2020. MAUI-seq: metabarcoding using amplicons with unique molecular identifiers to improve error correction. Mol Ecol Res 21, 703–720.
Froslev, T.G., Kjoller, R., Bruun, H.H., Ejrnaes, R., Brunbjerg, A.K., Pietroni, C., Hansen, A.J., 2017. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat. Commun. 8.
Ganley, A.R.D., Kobayashi, T., 2007. Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. Genome Res. 17, 184–191.
Gardes, M., Bruns, T.D., 1993. Its primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. Mol. Ecol. 2, 113–118.
Gardes, M., White, T.J., Fortin, J.A., Bruns, T.D., Taylor, J.W., 1991. Identification of indigenous and introduced symbiotic fungi in ectomycorrhizae by amplification of nuclear and mitochondrial ribosomal DNA. Can. J. Bot. 69, 180–190.
Garnica, S., Schon, M.E., Abarenkov, K., Riess, K., Liimatainen, K., Niskanen, T., Dima, B., Soop, K., Froslev, T.G., Jeppesen, T.S., Peintner, U., Kuhnert-Finkernagel, R., Brandrud, T.E., Saar, G., Oertel, B., Ammirati, J.F., 2016. Determining threshold values for barcoding fungi: lessons from Cortinarius (Basidiomycota), a highly diverse and widespread ectomycorrhizal genus. FEMS Microbiol. Ecol. 92.
Glassman, S.I., Martiny, J.B.H., 2018. Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. mSphere 3.
Gohl, D.M., Auch, B., Certano, A., LeFrancois, B., Bouevitch, A., Doukhanine, E., Fragel, C., Macklaim, J., Hollister, E., Garbe, J., Beckman, K.B., 2021. Dissecting and tuning primer editing by proofreading polymerases. Nucleic Acids Res. 49.
Guidot, A., Lumini, E., Debaud, J.C., Marmeisse, R., 1999. The nuclear ribosomal DNA intergenic spacer as a target sequence to study intraspecific diversity of the ectomycorrhizal basidiomycete Hebeloma cylindrosporum directly on Pinus root systems. Appl. Environ. Microbiol. 65, 903–909.
Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methe, B., DeSantis, T.Z., Petrosino, J.F., Knight, R., Birren, B.W., Consortium, H.M., 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res. 21, 494–504.
Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68, 87–112.
Hewitt, G.M., 2001. Speciation, hybrid zones and phylogeography - or seeing genes in space and time. Mol. Ecol. 10, 537–549.
Hibbett, D.S., Vilgalys, R., 1993. Phylogenetic-Relationships of lentinus (basidiomycotina) inferred from molecular and morphological characters. Syst. Bot. 18, 409–433.
Hoang, M.T.V., Irinyi, L., Chen, S.C.A., Sorrell, T.C., Meyer, W., Working, I.B.M.F., 2019. Dual DNA barcoding for the molecular identification of the agents of invasive fungal infections. Front. Microbiol. 10.
Horton, T.R., Bruns, T.D., 2001. The molecular revolution in ectomycorrhizal ecology: peeking into the black-box. Mol. Ecol. 10, 1855–1871.
Hughes, K.W., Petersen, R.H., Lickey, E.B., 2009. Using heterozygosity to estimate a percentage DNA sequence similarity for environmental species' delimitation across basidiomycete fungi. New Phytol. 182, 795–798.
James, T.Y., Moncalvo, J.M., Li, S., Vilgalys, R., 2001. Polymorphism at the ribosomal DNA spacers and its relation to breeding structure of the widespread mushroom Schizophyllum commune. Genetics 157, 149–161.
Jumpponen, A., Johnson, L.C., 2005. Can rDNA analyses of diverse fungal communities in soil and roots detect effects of environmental manipulations - a case study from tallgrass prairie. Mycologia 97, 1177–1194.
Jumpponen, A., Jones, K.L., 2009. Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate Quercus macrocarpa phyllosphere. New Phytol. 184, 438–448.
Kauserud, H., Schumacher, T., 2002. Population structure of the endangered wood decay fungus Phellinus nigrolimitatus (Basidiomycota). Can. J. Bot. 80, 597–606.
Kauserud, H., Schumacher, T., 2003a. Regional and local population structure of the pioneer wood-decay fungus Trichaptum abietinum. Mycologia 95, 416–425.
Kauserud, H., Schumacher, T., 2003b. Ribosomal DNA variation, recombination and inheritance in the basidiomycete Trichaptum abietinum: implications for reticulate evolution. Heredity 91, 163–172.
Kauserud, H., Hogberg, N., Knudsen, H., Elborne, S.A., Schumacher, T., 2004. Molecular phylogenetics suggest a North American link between the anthropogenic dry rot fungus Serpula lacrymans and its wild relative S-himantioides. Mol. Ecol. 13, 3137–3146.
Kauserud, H., Shalchian-Tabrizi, K., Decock, C., 2007a. Multilocus sequencing reveals multiple geographically structured lineages of Coniophora arida and C. olivacea (Boletales) in North America. Mycologia 99, 705–713.
Kauserud, H., Svegarden, I.B., Decock, C., Hallenberg, N., 2007b. Hybridization among cryptic species of the cellar fungus Coniophora puteana (Basidiomycota). Mol. Ecol. 16, 389–399.
Kauserud, H., Svegarden, I.B., Saetre, G.P., Knudsen, H., Stensrud, O., Schmidt, O., Doi, S., Sugiyama, T., Hogberg, N., 2007c. Asian origin and rapid global spread of the destructive dry rot fungus Serpula lacrymans. Mol. Ecol. 16, 3350–3360.
Koljalg, U., Dahlberg, A., Taylor, A.F.S., Larsson, E., Hallenberg, N., Stenlid, J., Larsson, K.H., Fransson, P.M., Karen, O., Jonsson, L., 2000. Diversity and abundance of resupinate thelephoroid fungi as ectomycorrhizal symbionts in Swedish boreal forests. Mol. Ecol. 9, 1985–1996.
Koljalg, U., Larsson, K.H., Abarenkov, K., Nilsson, R.H., Alexander, I.J., Eberhardt, U., Erland, S., Hoiland, K., Kjoller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A.F.S., Tedersoo, L., Vralstad, T., Ursing, B.M., 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol. 166, 1063–1068.
Koljalg, U., Nilsson, R.H., Abarenkov, K., Tedersoo, L., Taylor, A.F.S., Bahram, M., Bates, S.T., Bruns, T.D., Bengtsson-Palme, J., Callaghan, T.M., Douglas, B., Drenkhan, T., Eberhardt, U., Duenas, M., Grebenc, T., Griffith, G.W., Hartmann, M., Kirk, P.M., Kohout, P., Larsson, E., Lindahl, B.D., Luecking, R., Martin, M.P., Matheny, P.B., Nguyen, N.H., Niskanen, T., Oja, J., Peay, K.G., Peintner, U., Peterson, M., Poldmaa, K., Saag, L., Saar, I., Schuessler, A., Scott, J.A., Senes, C., Smith, M.E., Suija, A., Taylor, D.L., Telleria, M.T., Weiss, M., Larsson, K.H., 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22, 5271–5277.
Kopczynski, E.D., Bateson, M.M., Ward, D.M., 1994. Recognition of chimeric small-subunit ribosomal dnas composed of genes from uncultivated microorganisms. Appl. Environ. Microbiol. 60, 746–748.
Kunin, V., Engelbrektson, A., Ochman, H., Hugenholtz, P., 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. 12, 118–123.
Lekberg, Y., Gibbons, S.M., Rosendahl, S., 2014. Will different OTU delineation methods change interpretation of arbuscular mycorrhizal fungal community patterns? New Phytol. 202, 1101–1104.
Lindahl, B.D., Nilsson, R.H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjoller, R., Koljalg, U., Pennanen, T., Rosendahl, S., Stenlid, J., Kauserud, H., 2013. Fungal community analysis by high-throughput sequencing of amplified markers - a user's guide. New Phytol. 199, 288–299.

Lindner, D.L., Banik, M.T., 2011. Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus Laetiporus. Mycologia 103, 731–740.
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. Ecol. Evol. 3, 1751–1764.
Lipman, D.J., Pearson, W.R., 1985. Rapid and sensitive protein similarity searches. Science 227, 1435–1441.
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. Mol. Ecol. 28, 721–730.
Logares, R., Haverkamp, T.H.A., Kumar, S., Lanzen, A., Nederbragt, A.J., Quince, C., Kauserud, H., 2012. Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. J. Microbiol. Methods 91, 106–113.
Lucking, R., Aime, M.C., Robbertse, B., Miller, A.N., Ariyawansa, H.A., Aoki, T., Cardinali, G., Crous, P.W., Druzhinina, I.S., Geiser, D.M., Hawksworth, D.L., Hyde, K. D., Irinyi, L., Jeewon, R., Johnston, P.R., Kirk, P.M., Malosso, E., May, T.W., Meyer, W., Opik, M., Robert, V., Stadler, M., Thines, M., Vu, D., Yurkov, A.M., Zhang, N., Schoch, C.L., 2020. Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? IMA Fungus 11, 14.
Luo, M., Ji, Y., Warton, D., Yu, D.W., 2022. Extracting abundance information from DNA-based data. Mol Ecol Res 23, 174–189.
Mahe, F., Rognes, T., Quince, C., de Vargas, C., Dunthorn, M., 2014. Swarm: robust and fast clustering method for amplicon-based studies. PeerJ 2.
Mahe, F., Czech, L., Stamatakis, A., Quince, C., de Vargas, C., Dunthorn, M., Rognes, T., 2022. Swarm v3: towards tera-scale amplicon clustering. Bioinformatics 38, 267–269.
McKnight, D.T., Huerlimann, R., Bower, D.S., Schwarzkopf, L., Alford, R.A., Zenger, K.R., 2018. Methods for normalizing microbiome data: an ecological perspective. Methods Ecol. Evol. 10, 389–400.
McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput. Biol. 10, e1003531.
Mullis, K.B., 1990. The unusual origin of the polymerase chain reaction. Sci. Am. 262 (56–61), 64-55.
Mullis, K.B., Faloona, F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol. 155, 335–350.
Nilsson, R.H., Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K.H., Koljalg, U., 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS One 1.
Nilsson, R.H., Kristiansson, E., Ryberg, M., Hallenberg, N., Larsson, K.H., 2008. Intraspecific ITS variability in the kingdom fungi as expressed in the international sequence databases and its implications for molecular species identification. Evol. Bioinf. Online 4, 193–201.
Nilsson, R.H., Tedersoo, L., Ryberg, M., Kristiansson, E., Hartmann, M., Unterseher, M., Porter, T.M., Bengtsson-Palme, J., Walker, D.M., De Sousa, F., Gamper, H.A., Larsson, E., Larsson, K.H., Koljalg, U., Edgar, R.C., Abarenkov, K., 2015. A comprehensive, automatically updated fungal ITS sequence dataset for reference-based chimera control in environmental sequencing efforts. Microb. Environ. 30, 145–150.
Nilsson, R.H., Larsson, K.H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glockner, F.O., Tedersoo, L., Saar, I., Koljalg, U., Abarenkov, K., 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res. 47, D259–D264.
Nylund, J.E., Dahlberg, A., Högberg, N., Kårén, O., Grip, K., Jonsson, L., 1995. Methods for studying species composition of mycorrhizal fungal communities in ecological studies and environmental monitoring. In: Stocchi, V., Bonfante, P., Nuti, M. (Eds.), Biotechnology of Ectomycorrhizae. Plenum Press, New York, pp. 229–239.
O'Brien, H.E., Parrent, J.L., Jackson, J.A., Moncalvo, J.M., Vilgalys, R., 2005. Fungal community analysis by large-scale sequencing of environmental samples. Appl. Environ. Microbiol. 71, 5544–5550.
Öpik, M., Vanatoa, A., Vanatoa, A., Moora, M., Davison, J., Kalwij, J.M., Reier, Ü., Zobel, M., 2010. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). New Phytol. 188, 223–241.
O'Donnell, K., Cigelnik, E., 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus Fusarium are nonorthologous. Mol. Phylogenet. Evol. 7, 103–116.
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.
Peay, K.G., Kennedy, P.G., Bruns, T.D., 2008. Fungal community ecology: a hybrid beast with a molecular master. Bioscience 58, 799–810.
Potapov, V., Ong, J.L., 2017. Examining sources of error in PCR by single-molecule sequencing. PLoS One 12, e0169774.
Quince, C., Lanzen, A., Davenport, R.J., Turnbaugh, P.J., 2011. Removing noise from pyrosequenced amplicons. BMC Bioinf. 12.
Royo-Llonch, M., Sanchez, P., Ruiz-Gonzalez, C., Salazar, G., Pedros-Alio, C., Sebastian, M., Labadie, K., Paoli, L., Ibarbalz, F.M., Zinger, L., Churcheward, B., Chaffron, S., Eveillard, D., Karsenti, E., Sunagawa, S., Wincker, P., Karp-Boss, L., Bowler, C., Acinas, S.G., Coordinators, T.O., 2021. Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. Nature Microbiology 6, 1561.
Ryberg, M., 2015. Molecular operational taxonomic units as approximations of species in the light of evolutionary models and empirical data from Fungi. Mol. Ecol. 24, 5770–5777.
Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.
Schadt, C.W., Rosling, A., 2015. Comment on "Global diversity and geography of soil fungi". Science 348.
Schnell, I.B., Bohmann, K., Gilbert, M.T.P., 2015. Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. Mol Ecol Resour 15, 1289–1303.
Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K., Crous, P.W., Miller, A.N., Wingfield, M.J., Aime, M.C., An, K.D., Bai, F.Y., Barreto, R.W., Begerow, D., Bergeron, M.J., Blackwell, M., Boekhout, T., Bogale, M., Boonyuen, N., Burgaz, A.R., Buyck, B., Cai, L., Cai, Q., Cardinali, G., Chaverri, P., Coppins, B.J., Crespo, A., Cubas, P., Cummings, C., Damm, U., de Beer, Z.W., de Hoog, G.S., Del-Prado, R., Dentinger, B., Dieguez-Uribeondo, J., Divakar, P.K., Douglas, B., Duenas, M., Duong, T.A., Eberhardt, U., Edwards, J.E., Elshahed, M.S., Fliegerova, K., Furtado, M., Garcia, M. A., Ge, Z.W., Griffith, G.W., Griffiths, K., Groenewald, J.Z., Groenewald, M., Grube, M., Gryzenhout, M., Guo, L.D., Hagen, F., Hambleton, S., Hamelin, R.C., Hansen, K., Harrold, P., Heller, G., Herrera, G., Hirayama, K., Hirooka, Y., Ho, H.M., Hoffmann, K., Hofstetter, V., Hognabba, F., Hollingsworth, P.M., Hong, S.B., Hosaka, K., Houbraken, J., Hughes, K., Huhtinen, S., Hyde, K.D., James, T., Johnson, E.M., Johnson, J.E., Johnston, P.R., Jones, E.B., Kelly, L.J., Kirk, P.M., Knapp, D.G., Koljalg, U., Kovacs, G.M., Kurtzman, C.P., Landvik, S., Leavitt, S.D., Liggenstoffer, A.S., Liimatainen, K., Lombard, L., Luangsa-Ard, J.J., Lumbsch, H.T., Maganti, H., Maharachchikumbura, S.S., Martin, M.P., May, T.W., McTaggart, A.R., Methven, A.S., Meyer, W., Moncalvo, J.M., Mongkolsamrit, S., Nagy, L.G., Nilsson, R.H., Niskanen, T., Nyilasi, I., Okada, G., Okane, I., Olariaga, I., Otte, J., Papp, T., Park, D., Petkovits, T., Pino-Bodas, R., Quaedvlieg, W., Raja, H.A., Redecker, D., Rintoul, T.L., Ruibal, C., Sarmiento-Ramirez, J.M., Schmitt, I., Schussler, A., Shearer, C., Sotome, K., Stefani, F.O.P., Stenroos, S., Stielow, B., Stockinger, H., Suetrong, S., Suh, S.O., Sung, G.H., Suzuki, M., Tanaka, K., Tedersoo, L., Telleria, M.T., Tretter, E., Untereiner, W.A., Urbina, H., Vagvolgyi, C., Vialle, A., Vu, T.D., Walther, G., Wang, Q.M., Wang, Y., Weir, B.S., Weiss, M., White, M.M., Xu, J., Yahr, R., Yang, Z.L., Yurkov, A., Zamora, J.C., Zhang, N., Zhuang, W.Y., Schindel, D., Consortium, F.B., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. P Natl Acad Sci USA 109, 6241–6246.
Seierstad, K.S., Fossdal, R., Miettinen, O., Carlsen, T., Skrede, I., Kauserud, H., 2021. Contrasting genetic structuring in the closely related basidiomycetes Trichaptum abietinum and Trichaptum fuscoviolaceum (Hymenochaetales). Fungal Biol-Uk 125, 269–275.
Shelton, A.O., Gold, Z.J., Jensen, A.J., D'Agnese, E., Allan, E.A., Van Cise, A., Gallego, R., Ramón-Laca, A., Garber-Yonts, M., Parsons, K., Kelly, R.P., 2022. Toward quantitative metabarcoding. Ecology 104, e3906.
Shin, S., Lee, T.K., Han, J.M., Park, J., 2014. Regional effects on chimera formation in 454 pyrosequenced amplicons from a mock community. J. Microbiol. 52, 566–573.
Simon, U.K., Weiss, M., 2008. Intragenomic variation of fungal ribosomal genes is higher than previously thought. Mol. Biol. Evol. 25, 2251–2254.
Simon, L., Bousquet, J., Levesque, R.C., Lalonde, M., 1993. Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. Nature 363, 67–69.
Skaven Seierstad, K., Carlsen, T., Saetre, G.P., Miettinen, O., Hellik Hofton, T., Kauserud, H., 2013. A phylogeographic survey of a circumboreal polypore indicates introgression among ecologically differentiated cryptic lineages. Fungal Ecol 6, 119–128.
Sokal, R.R., 1963. The principles and practice of numerical taxonomy. Taxon 12, 190–199.
Somervuo, P., Koskela, S., Pennanen, J., Nilsson, R.H., Ovaskainen, O., 2016. Unbiased probabilistic taxonomic classification for DNA barcoding. Bioinformatics 32, 2920–2927.
Stammler, F., Glasner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P.J., Gessner, A., Spang, R., 2016. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. Microbiome 4.
Stielow, J.B., Levesque, C.A., Seifert, K.A., Meyer, W., Irinyi, L., Smits, D., Renfurm, R., Verkley, G.J.M., Groenewald, M., Chaduli, D., Lomascolo, A., Welti, S., Lesage-Meessen, L., Favel, A., Al-Hatmi, A.M.S., Damm, U., Yilmaz, N., Houbraken, J., Lombard, L., Quaedvlieg, W., Binder, M., Vaas, L.A.I., Vu, D., Yurkov, A., Begerow, D., Roehl, O., Guerreiro, M., Fonseca, A., Samerpitak, K., van Diepeningen, A.D., Dolatabadi, S., Moreno, L.F., Casaregola, S., Mallet, S., Jacques, N., Roscini, L., Egidi, E., Bizet, C., Garcia-Hermoso, D., Martin, M.P., Deng, S., Groenewald, J.Z., Boekhout, T., de Beer, Z.W., Barnes, I., Duong, T.A., Wingfield, M.J., de Hoog, G.S., Crous, P.W., Lewis, C.T., Hambleton, S., Moussa, T.A. A., Al-Zahrani, H.S., Almaghrabi, O.A., Louis-Seize, G., Assabgui, R., McCormick, W., Omer, G., Dukik, K., Cardinali, G., Eberhardt, U., de Vries, M., Robert, V., 2015. One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. Persoonia 35, 242–263.
Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. P Natl Acad Sci USA 114, 1607–1612.
Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7, 453–464.
Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012a. Environmental DNA. Mol. Ecol. 21, 1789–1793.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012b. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21, 2045–2050.
Taberlet, P., Bonin, A., Zinger, L., Coissac, E., 2018. Environmental DNA: for Biodiversity Research and Monitoring. Oxford University Press, New York.
Taylor, J.W., Jacobson, D.J., Kroken, S., Kasuga, T., Geiser, D.M., Hibbett, D.S., Fisher, M.C., 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31, 21–32.
Taylor, J.W., Turner, E., Townsend, J.P., Dettman, J.R., Jacobson, D., 2006. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. Philos T R Soc B 361, 1947–1963.
Tedersoo, L., Lindahl, B., 2016. Fungal identification biases in microbiome projects. Env Microbiol Rep 8, 774–779.
Tedersoo, L., Albertsen, M., Anslan, S., Callahan, B., 2021. Perspectives and benefits of high-throughput long-read sequencing in microbial ecology. Appl. Environ. Microbiol. 87.
Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R.H., Kennedy, P.G., Yang, T., Anslan, S., Mikryukov, V., 2022. Best practices in metabarcoding of fungi: from experimental design to results. Mol. Ecol. 31, 2769–2795.
Vandenkoornhuyse, P., Husband, R., Daniell, T.J., Watson, I.J., Duck, J.M., Fitter, A.H., Young, J.P.W., 2002. Arbuscular mycorrhizal community composition associated with two plant species in a grassland ecosystem. Mol. Ecol. 11, 1555–1564.
Vetrovsky, T., Kolarik, M., Zifcakova, L., Zelenka, T., Baldrian, P., 2016. The rpb2 gene represents a viable alternative molecular marker for the analysis of environmental fungal communities. Mol Ecol Resour 16, 388–401.
Vetrovsky, T., Morais, D., Kohout, P., Lepinay, C., Algora, C., Holla, S.A., Bahnmann, B. D., Bilohneda, K., Brabcova, V., D'Alo, F., Human, Z.R., Jomura, M., Kolarik, M., Kvasnickova, J., Llado, S., Lopez-Mondejar, R., Martinovic, T., Masinova, T., Meszarosova, L., Michalcikova, L., Michalova, T., Mundra, S., Navratilova, D., Odriozola, I., Piche-Choquette, S., Stursova, M., Svec, K., Tlaskal, V., Urbanova, M., Vlk, L., Voriskova, J., Zifcakova, L., Baldrian, P., 2020. GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. Sci. Data 7.
Vilgalys, R., Hester, M., 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several cryptococcus species. J. Bacteriol. 172, 4238–4246.
Vralstad, T., Myhre, E., Schumacher, T., 2002. Molecular diversity and phylogenetic affinities of symbiotic root-associated ascomycetes of the Helotiales in burnt and metal polluted habitats. New Phytol. 155, 131–148.
Vu, D., Groenewald, M., de Vries, M., Gehrmann, T., Stielow, B., Eberhardt, U., Al-Hatmi, A., Groenewald, J.Z., Cardinali, G., Houbraken, J., Boekhout, T., Crous, P.W., Robert, V., Verkley, G.J.M., 2019. Large-scale generation and analysis of filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal species and higher taxon delimitation. Stud. Mycol. 135–154.
Wang, G.C.Y., Wang, Y., 1996. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiol. 142, 1107–1114.
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. Appl. Environ. Microbiol. 73, 5261–5267.
Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E.R., Knight, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5, 27.
White, T.J., Bruns, T., Lee, S., Taylor, J.W., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press Incorporation, New York, pp. 315–322.
Wurzbacher, C., Larsson, E., Bengtsson-Palme, J., Van den Wyngaert, S., Svantesson, S., Kristiansson, E., Kagami, M., Nilsson, R.H., 2019. Introducing ribosomal tandem repeat barcoding for fungi. Mol Ecol Resour 19, 118–127.
Yang, R.H., Su, J.H., Shang, J.J., Wu, Y.Y., Li, Y., Bao, D.P., Yao, Y.J., 2018. Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. PLoS One 13.
Zinger, L., Bonin, A., Alsos, I.G., Balint, M., Bik, H., Boyer, F., Chariton, A.A., Creer, S., Coissac, E., Deagle, B.E., Barba, M., Dickie, I.A., Dumbrell, A.J., Ficetola, G.F., Fierer, N., Fumagalli, L., Gilbert, M.T.P., Jarman, S., Jumpponen, A., Kauserud, H., Orlando, L., Pansu, J., Pawlowski, J., Tedersoo, L., Thomsen, P.F., Willerslev, E., Taberlet, P., 2019. DNA metabarcoding-Need for robust experimental designs to draw sound ecological conclusions. Mol. Ecol. 28, 1857–1862.