Case Report Writing
Sample Work

Effective methods for metabarcoding fungi: from experimental design to outcomes

Abstract

The advancement of high-throughput Sequencing (HTS) technology has dramatically improved our ability to detect fungi and reveal their ecological research importance in a wide range of habitats.

Overview

From experimental design to molecular and computational analyses, we present an overview of current best practices in metabarcoding investigation of fungal communities. We show that operational taxonomic units (OTUs) outperform amplified sequence variants (ASVs) in recovering fungal diversity, especially for lengthy markers, by reanalyzing published data collection sets. Furthermore, as compared to the ITS2 subregion, examination of the whole ITS region allows for more precise taxonomic placement of fungus and other eukaryotes. Finally, we show that specialized strategies for analyzing compositional data yield more trustworthy estimations of community structure alterations.

Conclusion:

We conclude that fungus metabarcoding investigations are especially promising for integrating fungi into the complete microbiome and larger systematic ecosystem functioning environment, recovering new fungal lineages and ancient species, and barcoding old specimens, including type material.

Keywords:

Biodiversity, bioinformatics, community ecology, experiment setup, molecular identification, statistical analyses

INTRODUCTION

Fungi have typically been recognized using macromorphological and micromorphological characteristics of fruiting body specimens or pure cultures. The development of molecular methods in the late 1980s constituted a substantial advancement in fungus identification. PCR amplification and Sanger sequencing of the nuclear 18S (SSU) and 28S (LSU) ribosomal rRNA genes, as well as the nuclear ribosomal internal transcribed spacer (ITS) region from fungal tissue (e.g., lichen thalli, lesions in plant and animal tissue, cultures from environmental samples, and ectomycorrhizal root tips) quickly became popular and provided unprecedented taxonomic resolution. Common applications included species- and genus-level identification, cryptic species analysis, and phylogenetic study of important fungal clades, as well as the kingdom of Fungi as a whole (Gherbawy & Voigt, 2010). Later, by integrating a cloning phase of amplicons prior to Sequencing, it became able to identify numerous fungi from more varied substrates, including soil, plant roots, and water. However, these investigations often used tens to low hundreds of readings rather than the thousands needed to quantify fungal diversity in soils accurately (Taylor et al., 2014). As a result, sequences and operational taxonomic units (OTUs) were often processed manually or with specialized software, with no pressing need for bioinformatics tools.

Bioinformatic platforms and analytical resources grew in tandem with the fast growth of HTS technologies to meet the computing demands imposed by enormous data sets. Metabarcoding approaches have been extensively reviewed in several recent studies with a focus on their conceptual foundation (Taberlet et al., 2018), pathogenic organisms (Piombo et al., 2021; Tedersoo et al., 2019), applications in mycology (Lindahl et al., 2013; Nilsson et al., 2018), eukaryotes more broadly (Ruppert et al., 2019 We present a survey of various approaches and recommend best practices for developing and carrying out metabarcoding investigations.

2.DEVELOPING A METABARCODING STUDY

Researchers should first examine a good methodological experimental design - either observational, experimental, or combination - encompassing technological, analytical, personnel-related, and budgetary needs before testing scientific ideas. Experiment designs with wide representativeness (e.g., geographical and ecological scope) and replication independence (i.e., no spatiotemporal autocorrelation) are strongly advised (Gotelli & Ellison, 2013; Zinger, Bonin, et al., 2019). Indeed, metabarcoding investigations are similar to classic ecological studies in that the number and distribution of research sites must be suitably determined based on the initial query (Dickie et al., 2018). Furthermore, metabarcoding investigations need an ideal number of local biological replicates, which may be calculated using variance reported in earlier research (Alteio et al., 2021) or pilot trials.

Pooling statistically nonindependent subsamples is a popular method for improving sample representativeness at a low cost. The number and spatial distance of subsamples may be crucial in providing a realistic representation of microbial variety in varied ecosystems; less inclusive subsampling strategies are likely to underestimate diversity (Figure Box 1). The number of subsamples to pool is determined by the study topic and the size of the area, with 7-25 being the most common (Schwarzenbach et al., 2007). Both physical and analytical pooling enhance evaluations of soil fungus richness and composition (Schwarzenbach et al., 2007; Song et al., 2015) and minimize estimated variation (Dickie et al., 2018). Pooling of physical samples, on the other hand, may result in loss. These findings might be useful for fungal groups with low DNA content and rRNA copy counts, such as Glomeromycota and unicellular taxa. The pooling effect is likely to be influenced by habitat heterogeneity, such as pH, organic matter content, salinity, and plant species present - all of which are known to influence fungal composition in various environments (Amend et al., 2019; Grossart et al., 2019; Nilsson et al., 2019; U'Ren et al., 2019). As a result, pooling samples with possibly varied microbial compositions (for example, leaves from various plant species) is not advised. Pooling does not perform ideally in circumstances when the samples include varying quantities of DNA, and the low-DNA samples contain distinct, rare species. Pooled samples need further analysis due to their increased overall richness. Furthermore, pooling is unsuitable for analyzing biotic interactions in co-occurrence analysis (Bahram et al., 2014). When individual samples cannot be used as independent replicates (local- or landscape-scale spatial autocorrelation), for example, for regional- to global-scale analyses, pooling them at the site level (during the DNA extraction, PCR, library preparation, or sequence data) may be the most useful.

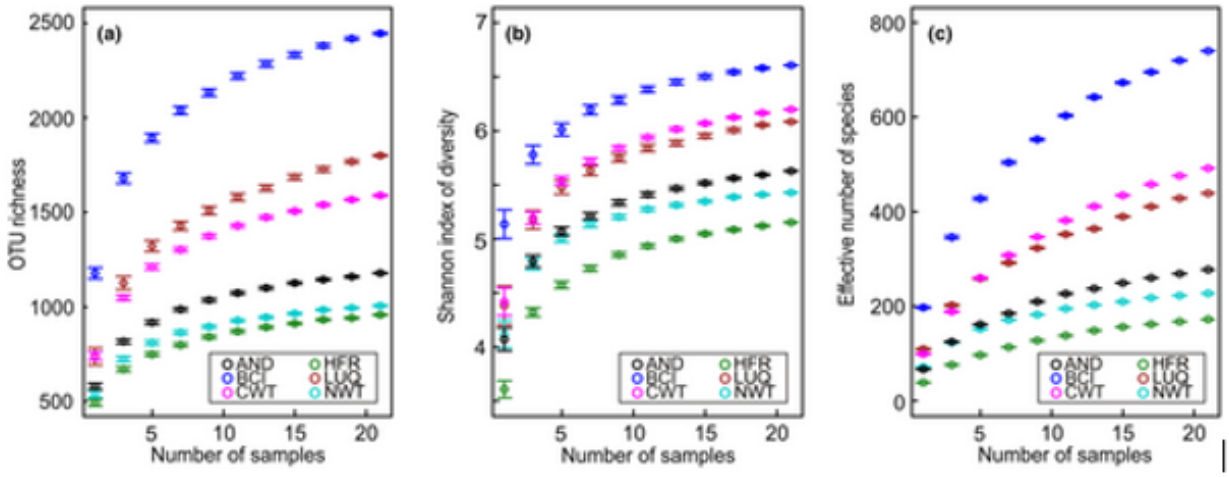

Figure: Potential underestimation of biodiversity and high variance at the low number of (sub)samples (Leho Tedersoo et al. 2022)

3.SAMPLING AND RETAINING

To avoid contaminating samples with skin or forward bacteria, we highly advise wearing disposable gloves when sampling. To reduce the risk of cross-contamination between independent samples in the field, sampling tools should be replaced or sterilized between sampling events using oxidizing agents (e.g., bleach; Fischer et al., 2016), DNase solutions, or flame, rather than just alcohols, which do not denature DNA. Paper bags (leaf and fruiting body material), plastic bags (roots, soil, and sediments), or screw-cap jars (soil, water, and sediments) should be used to collect samples. To minimize contamination from rain and wet hands, it is best to collect samples in the field during dry weather. It is also advised to incorporate field controls (e.g., empty tubes left open at the site or extraction of sample storage buffer) in the experiment to enable a posteriori removal of site or sampling material contamination (Zinger, Bonin, et al., 2019). Finally, it is critical to minimize biological activity inside samples post-harvest (i.e., the growth of fast-growing molds), which may be accomplished by keeping the samples cool throughout transport.

4.ANALYSIS OF MOLECULES

Prior to DNA extraction, the material must be homogenized by bead beating in microcentrifuge tubes, using a mortar and pestle, micro pellets, or knife mill for small sample numbers. The appropriate amount of material should be weighted into the DNA extraction tube, with the remainder saved for backup purposes, such as stable isotopes or chemical studies. Reaching the maximum capacity of the DNA extraction kit is normally undesired because different types of materials (e.g., peat soils, dead wood, debris-rich sediments, and fleshy plant tissues) may absorb the liquid or inhibitors may be coextracted. There are only minor differences in perceived richness when using DNA extracts from 0.25, 1 or 10 g material for well-homogenized soil samples (Song et al., 2015), but increasing the volume through replicate extractions or more material using "maxi" kits provides more reproducible estimates (Dickie et al., 2018). To minimize cross-contamination and air contamination by amplicons, weigh and extract DNA in a room separate from the PCR laboratory using a dedicated laminar flow. By analyzing blank DNA extraction controls, such possible contaminants may be recognized and deleted in subsequent analyses.

Ecologists usually employ primers created decades ago for Sanger sequencing investigations for metabarcoding (Figure 1; White et al., 1990). These initial primers are ineffective for the various fungal groups with one or more primer-template mismatches. They can be enhanced by including degenerate locations to reduce primer bias and boost quantitative performance (Tedersoo & Lindahl, 2016; Pinol et al., 2019). However, because not all primer variations match templates, numerous degeneracies may necessitate changing the 1:1 ratio of primers and further PCR cycles. Because of multiple key mismatches in some types of moulds and suspected animal diseases, the widely used fungus-specific forward primer ITS1F is particularly problematic (Tedersoo & Lindahl, 2016). Researchers should also consider the prevalence of an intron at the end of the 18S rRNA gene, which prohibits the Sequencing of taxa containing this intron (Figure 1). To achieve the best performance, it may be necessary to couple primers with identical melting temperatures.

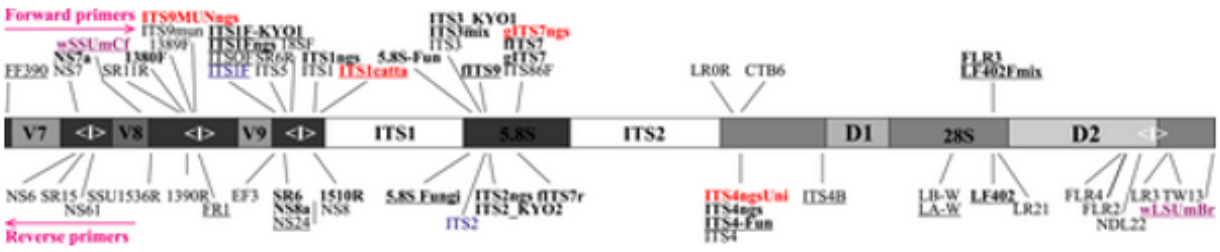

Figure1 : Primer map of the rRNA operon internal transcribed spacer (ITS) region (Leho Tedersoo et al. 2022)

5.ANALYSIS OF BIOINFORMATICS DATA

Sequencing instruments' raw output is transformed to fastq format, which is compatible with all main quality-filtering programs. Because most bioinformatic systems were designed for bacterial 16S data, their ability to handle fungal ITS sequences, which often cannot be properly matched beyond the genus level, varies substantially (Anslan et al., 2018). According to citations, the most popular platforms are QIIME2, Mother, PIPITS, SEED2, SCATA, and PipeCraft—the most popular and newly created platforms' features, as well as their advantages and disadvantages.

Index-switches (also known as tag-switches, index-jumping, and index cross-talk) are the most damaging phenomenon in HTS data, causing technical cross-contamination among samples and potentially blurring patterns in host specificity, taxon networks, and biogeographical patterns (Calderon-Sanou et al., 2020; Carlsen et al., 2012). Index changes are known from all sequencing platforms and occur during PCR, T4 blunt-ending, and cross-pairing of amplicons from distinct libraries (Care & Bohmann, 2020; Schnell et al., 2015). Although careful sample indexing alleviates this problem, around 0.01-0.1% of evident transitions will remain. The UNCROSS method (Edgar, 2018), the unspread Python script (Larsson et al., 2018), tracking nonbiological spike-ins (Palmer et al., 2018), or a positive control sample may all be used to examine index switches. Index-switch rates may be determined based on the distribution of spike-ins or positive controls in biological samples and vice versa. Sequence abundances that are less than the index-switching threshold are set to zero. Index switches and other contaminants should be examined independently for each sequencing collection. To quantify the fraction of index-switch-like artifacts, it is also beneficial to estimate the incidence of taxa in pseudosamples represented by unused indexes (see Taberlet et al., 2018; Zinger et al., 2021).

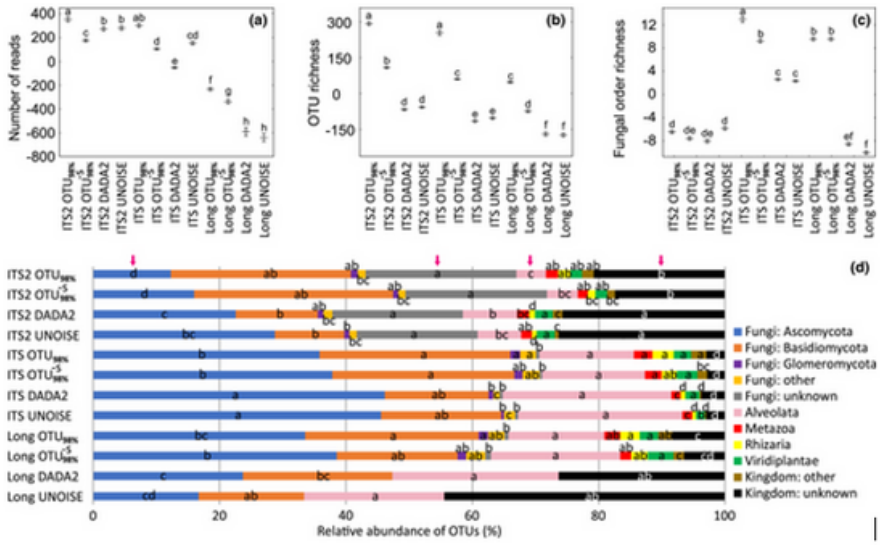

Figure2 : Comparison of traditional OTU-based (OTU−s, singletons removed) and ESV-based approaches (DADA2 and UNOISE) (Leho Tedersoo et 2022)

6.ANALYZING STATISTICAL DATA

Because macroscopic members of these groups may be easily differentiated by morphology, most mycologists, along with plant and animal ecologists, aim to approximate the species level in biodiversity assessments (Antich et al., 2021; Stajich et al., 2009). In contrast, it is far more difficult to distinguish species in bacteria. Traditionally, OTU richness, Shannon and Simpson indices, and the corresponding effective species numbers (Hill numbers; Alberdi & Gilbert, 2019; Chao et al., 2014) have been used to study the molecular diversity of organisms. Diversity indices only seldom correlate with sequencing depth because they reduce the impact of uncommon OTUs. OTU richness exhibits a cumulative effect with sequencing depth that is especially noticeable in pooled, heterogeneous data. Unless rarefaction is used, it is critical to add square-root or log-transformed sequencing depth as a covariate (whichever is most informative). When compared to untransformed data, rarefied data, and diversity indices in large data sets, log-transformation of fungal OTU richness accounting for sequencing depth gives better-explained statistical models (Tedersoo et al., 2022). Even though they are widely used, we do not suggest OTU richness extrapolations (e.g., Chao1 and ACE) since they rely on the number of the rarest OTUs, which is frequently artefactual (Balint et al., 2016; Bunge et al., 2014).

Adding phylogenetic information to taxonomic composition decreases the influence of any residual PCR/sequencing mistakes in the data (Taberlet et al., 2018) and eliminates the ambiguity around OTU estimates (Washburne et al., 2018). Much above the genus level, the ITS region is not susceptible to strong multiple alignments and phylogenetic reconstruction. Therefore, phylogenetic measures require inferring phylogenetic distance matrices that may rely on ultrametric trees of conserved gene(s) (Davison et al., 2015; Horn et al., 2014), grafting phylogenies (Fouquier et al., 2016) or mapping of OTUs to distance-weighted phylogenies (Perez-Izquierdo et al., 2019) or hierarchical taxonomic trees (Chalmandrier et al., 2019; Tedersoo, Sánchez-Ramírez, et al., 2018). For species-level identification, we propose examining the ITS region and flanking, phylogenetically relevant 18S or 28S rRNA genes. Standardized phylogenetic diversity (PD; averaged unique branch length), mean phylogenetic distance (MPD), UniFrac distance, and mean nearest taxon distance (MNTD) can all be used to investigate shifts in phylogenetic diversity. Analyses of plant-fungal interactions (Chalmandrier et al., 2019), fungal community assembly processes (Roy et al., 2019), and phylogeographic patterns (Turon et al., 2020) may benefit from testing phylogenetic conservatism, overdispersion, and turnover across phylogenetic scales (Tucker et al., 2017). These and other community turnover indices may be estimated using phylocom (Webb et al., 2008), the R packages picante (Kembel et al., 2010), S.phylomaker (Qian & Jin, 2016), PhyloMeasures (Tsirogiannis & Sandel, 2016), and other open-access scripts (Chalmandrier et al., 2019).

FUTURE PERSPECTIVES AND CONCLUSIONS

HTS studies have lately pushed the boundaries of several disciplines of mycology. Metabarcoding investigations of labelled substrates employing SIP (Hannula et al., 2017) or combined with metatranscriptomics (ifáková et al., 2016) have identified functionally active fungi and their activity in situ. Parallel research on fungi, bacteria, and protists has offered insight into antagonistic interactions (Bahram et al., 2018; Bork et al., 2015), the structure of the micro- and mycobiome web (Tipton et al., 2018), and community assembly processes (Zinger, Taberlet, et al., 2019). Several groups of hitherto undescribed (or unsequenced) order- and class-level fungal lineages have been discovered using HTS data (Tedersoo, Anslan, Bahram, Kljalg, et al., 2020; Zhang et al., 2021). HTS readings provide material for the development of taxon-specific primers and probes for cell visualization (Chambouvet). This may be especially important for fungal species that are difficult to cultivate and do not produce fruit bodies (Lücking et al., 2021). Metabarcoding of small segments also offers insight into ancient DNA, perhaps including ancient fungi (Balint et al., 2018; Talas et al., 2021). Similarly, metabarcoding approaches are useful in generating DNA barcodes from polyploid organisms (Maeda et al., 2018), organisms with multiple haplotypes (Runnel et al., 2022), or organisms represented by century-old specimens containing valuable type material where extra care is required to identify and dismiss air-borne contaminants (Forin et al., 2018).

Fungal diversity has been studied using short-read metabarcoding in virtually all ecosystems on Earth, including severe settings (Nilsson et al., 2018). We expect that, with the help of expanding reference databases, investigations incorporating fungal taxonomic, phylogenetic, and functional composition will flourish because many key elements of evolutionary and functional (e.g., trait-based) ecology are still little understood. To confirm the findings and establish causality, these investigations should be reinforced by rigorous experiments. It seems inexcusable that research on one of the most important groups of nutrient cyclers, fungi, should be based on anything less than the greatest and most up-to-date methodological principles, and we really hope that our study has helped to achieve that goal.

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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