# Understudied, underrepresented, and unknown: methodological biases that limit detection of early diverging fungi from environmental samples

Nicole Reynolds[1], Michelle Jusino[2], Jason Stajich[3], and Matthew Smith[1]

[1]University of Florida
[2]Williams and Mary College
[3]University of California Riverside

April 14, 2021

**Abstract**

Metabarcoding is an important tool for understanding fungal communities. The internal transcribed spacer (ITS) rDNA is the accepted fungal barcode but has known problems. The large subunit (LSU) rDNA has also been used to investigate fungal communities but available LSU metabarcoding primers were mostly designed to target Dikarya (Ascomycota + Basidiomycota) with little attention to early diverging fungi (EDF). However, evidence from multiple studies suggests that EDF comprise a large portion of unknown diversity in community sampling. Here we investigate how DNA marker choice and methodological biases impact recovery of EDF from environmental samples. We focused on one EDF lineage, Zoopagomycota, as an example. We evaluated three primer sets (ITS1F/ITS2, LROR/LR3, and LR3 paired with new primer LR22F) to amplify and sequence a Zoopagomycota mock community and a set of 146 environmental samples with Illumina MiSeq. We compared two taxonomy assignment methods and created an LSU reference database compatible with AMPtk software. The two taxonomy assignment methods recovered strikingly different communities of fungi and EDF. Target fragment length variation exacerbated PCR amplification biases and influenced downstream taxonomic assignments, but this effect was greater for EDF than Dikarya. To improve identification of LSU amplicons we performed phylogenetic reconstruction and illustrate the advantages of this critical tool for investigating identified and unidentified sequences. Our results suggest much of the EDF community may be missed or misidentified with "standard" metabarcoding approaches and modified techniques are needed to understand the role of these taxa in a broader ecological context.

**Title: Understudied, underrepresented, and unknown: methodological biases that limit detection of early diverging fungi from environmental samples**

**Short title: Metabarcoding biases limit detection of EDF**

Nicole K. Reynolds[1*], Michelle A. Jusino[2], Jason E. Stajich[3], Matthew E. Smith[1]

[1] *University of Florida, Department of Plant Pathology, Gainesville, Florida 32611*

[2] *Williams and Mary College, Department of Biology, Williamsburg, Virginia 23185*

[3] *Department of Plant Pathology & Microbiology and Institute for Integrative Genome Biology, University of California–Riverside, Riverside, California 92521*

**Corresponding author: nicolereynolds1@ufl.edu*

**Abstract**

Metabarcoding is an important tool for understanding fungal communities. The internal transcribed spacer (ITS) rDNA is the accepted fungal barcode but has known problems. The large subunit (LSU) rDNA has also been used to investigate fungal communities but available LSU metabarcoding primers were mostly designed to target Dikarya (Ascomycota + Basidiomycota) with little attention to early diverging fungi (EDF). However, evidence from multiple studies suggests that EDF comprise a large portion of unknown diversity in community sampling. Here we investigate how DNA marker choice and methodological biases impact recovery of EDF from environmental samples. We focused on one EDF lineage, Zoopagomycota, as an example. We evaluated three primer sets (ITS1F/ITS2, LROR/LR3, and LR3 paired with new primer LR22F) to amplify and sequence a Zoopagomycota mock community and a set of 146 environmental samples with Illumina MiSeq. We compared two taxonomy assignment methods and created an LSU reference database compatible with AMPtk software. The two taxonomy assignment methods recovered strikingly different communities of fungi and EDF. Target fragment length variation exacerbated PCR amplification biases and influenced downstream taxonomic assignments, but this effect was greater for EDF than Dikarya. To improve identification of LSU amplicons we performed phylogenetic reconstruction and illustrate the advantages of this critical tool for investigating identified and unidentified sequences. Our results suggest much of the EDF community may be missed or misidentified with "standard" metabarcoding approaches and modified techniques are needed to understand the role of these taxa in a broader ecological context.

**Key Words**

Illumina MiSeq, ITS, LSU, Metabarcoding, Mock community, Zoopagomycota

## Introduction

Metabarcoding of fungal communities using high-throughput technologies is a powerful tool for investigating fungal ecology. The internal transcribed spacer (ITS) region of the rDNA operon has been used extensively as the DNA barcode for fungi, particularly in environmental sequencing studies (Gardes & Bruns, 1996; Schoch et al., 2012). Because the ITS region has been used as a barcode for approximately two decades, the bioinformatic tools for processing amplicon data are well developed and the ITS reference databases (UNITE and INSD) are consistently curated and updated (Abarenkov et al., 2020; Nilsson et al., 2019). However, even with the widespread usage of these databases there are problems, including poor representation of some taxonomic groups, low quality sequences, and incorrect taxonomic annotation (Abarenkov et al., 2018; Hofstetter et al., 2019; Nilsson et al., 2012). These problems mean that results from taxonomic assignments of operational taxonomic units (OTUs) must be interpreted with caution (Nilsson et al., 2006; Yahr et al., 2016). Despite its widespread use, the ITS is not a suitable barcode for all taxonomic groups. There are known issues with sequencing ITS in some fungal lineages, including high variability in ITS length between groups (resulting in favored PCR and sequencing for shorter fragments) (Castaño et al., 2020; Engelbrektson et al., 2010; Manter & Vivanco, 2007), lack of interspecific discrimination (and therefore inability to use the marker for species-level determinations) (e.g. Gazis et al., 2011), interspecific rDNA copy number variation (Lindner & Banik, 2011), and primer biases which exclude some groups (Bellemain et al., 2010; Li et al., 2020; Tedersoo & Lindahl, 2016). For example, among arbuscular mycorrhizal fungi (AMF) the ITS is hypervariable and has high intraspecific and intra-spore variation compared to the small (SSU) and large (LSU) rDNA subunits (Egan et al., 2018; Thiéry et al., 2012). One study found up to 6% divergence among sequences from a single spore (Lloyd-MacGilp et al., 1996). A recent study also found wide rDNA copy number variation across kingdom Fungi that was uncorrelated with trophic mode (Lofgren et al., 2019), making such variation unpredictable in environmental samples. Interspecific rDNA variation can lead to the formation of multiple OTUs derived from a single individual and individuals with more rDNA copies could potentially dominate during PCR amplification and sequencing from mixed templates. Among early diverging fungal (EDF) lineages, direct comparisons of markers for metabarcoding have not been performed for many groups. An exception are the AMF for which the SSU, LSU, and ITS have been evaluated and some combination of two markers is commonly used (Hart et al., 2015; Öpik et al., 2014). Additionally, the LSU was suggested to perform better than ITS as a barcode for EDF due to greater PCR success and a larger barcode gap (i.e. difference between inter- and intraspecific variation) than SSU (Schoch et al., 2012).

Another critical advantage of the LSU and SSU is the ability to perform phylogenetic reconstruction with the OTUs to provide preliminary placement of unidentified sequences.

The majority of fungal metabarcoding studies have focused on the Dikarya. In contrast, EDF such as chytrids and zygomycetes are often overlooked or ignored in environmental sequencing studies. EDF are also generally missing or underrepresented during attempts to develop "universal" fungal primers for metabarcoding. For example, mock communities used to validate primer performance often contain none or a few EDF, despite the fact that these fungi often have highly divergent target sequences relative to Dikarya (e.g. Ihrmark et al., 2012; Pérez-Izquierdo et al., 2017; Stielow et al., 2015; Tedersoo et al., 2015). Early diverging lineages are among the least studied fungi and are generally challenging to collect and manipulate in the lab. Many EDF are not culturable using standard techniques, are obligate symbionts, and have limited or no sequence data available (e.g. Benny et al. 2016; Corsaro et al. 2014, 2017; Lazarus & James, 2015; Letcher & Powell, 2019; Malar et al. 2021). Available data suggest that a large portion of undescribed taxa belong to EDF lineages (Tedersoo et al. 2017, 2020; Torres-Cruz et al., 2017; Walsh et al., 2020). Metabarcoding methods can provide an essential tool for learning more about these "dark matter fungi" (Grossart et al., 2016), especially if taxonomy can be reliably assigned to the OTUs. Among the most understudied groups of EDF is the Zoopagomycota. The placement of Zoopagomycota is still unresolved but phylogenomic studies indicate this lineage is either sister to all other terrestrial fungi (Dikarya + Mucoromycota – Spatafora et al., 2016) or sister to the Mucoromycota (Li et al., 2021). This phylum is ecologically diverse and includes fungal parasites (mycoparasites) as well as parasites of small animals. Specialized enrichment methods indicate that some taxa are diverse and widespread in soils, leaf litter, and dung (e.g. Benjamin, 1958; Benny et al., 2016; Drechsler, 1938; Duddington, 1955). Despite these findings, Zoopagomycota species are absent or found in low abundance in most metabarcoding studies (Lazarus et al., 2017; Reynolds et al., 2019).

We focused on some species of Zoopagomycota as a test case for evaluating methodological biases because: 1) these fungi are routinely found in soil during culture-based studies (e.g. Benjamin 1958; Benny et al., 2016) and yet they are not readily detected with metabarcoding, 2) there are limited reference sequence data available, 3) among species for which sequence data are available there is large ITS sequence length variation (Lazarus et al., 2017; Reynolds et al., 2019), and 4) they are difficult to isolate, culture, and work with in the lab, making environmental sampling a particularly important investigative tool. Out of the three subphyla (Entomophthoromycotina, Kickxellomycotina, and Zoopagomycotina) we focused on taxa that have been isolated from soil samples. This includes members of the Zoopagomycotina and Kickxellomycotina that have been regularly isolated from our study sites and for which we have a large culture collection. Zoopagomycotina species are mycoparasites that primarily attack host fungi in the Mucoromycota (and sometimes Ascomycota) or are parasites of microinvertebrates (e.g. amoebae, nematodes, rotifers) (Zoopagales). The Kickxellomycotina is the most diverse subphylum and includes mycoparasites (Dimargaritales) that are parasitic on Mucoromycota and Ascomycota, commensalistic arthropod gut-dwelling species (Asellariales, Harpellales, Orphellales), and putative saprotrophs (Kickxellales). We collected soil (Dimargaritales, Kickxellales, Zoopagales), freshwater sediments and water (Harpellales, Orphellales), and microinvertebrate (Zoopagales) samples from multiple sites in California (CA) and Florida (FL), two geographically distant states with divergent climate and soils. Both locations have been heavily sampled for Zoopagomycota fungi using selective culturing methods (Benjamin 1958, 1959, 1961; Benny et al., 2016; Lazarus et al., 2017; Reynolds et al., 2019). We did not include Entomophthoromycotina because they are obligate arthropod pathogens and likely to be rare or absent from many environmental samples.

In order to detect fungal communities, we compared the ITS1 marker region to two different regions of the LSU rDNA, using two primer pairs for LSU (LROR/LR3 and LR3 paired with the newly designed primer LR22F). Because relatively few studies have used LSU as a sole marker for profiling fungal communities, we also compared the efficacy of two methods for LSU OTU taxonomy assignment: the RDP Naïve Bayesian Classifier (hereafter RDP classifier) (Wang et al., 2007) and UTAX (Edgar, 2010) to search a manually curated reference database. We created the new LSU database by combining the RDP (Cole et al., 2014) and SILVA (Quast et al., 2013) reference databases along with additional sequences from GenBank and our lab. Finally, we used a subset of LSU OTUs generated from each primer pair that were identified as

Zoopagomycota or only to kingdom Fungi and included them in phylogenetic reconstructions. Analyzing amplicons in this way allowed us to validate the taxonomic assignments made by our pipeline as well as identify putative errors, a step that is not possible with ITS data. We created a Zoopagomycota mock community that was evaluated alongside the environmental samples to investigate how the metabarcoding pipeline affected that group specifically. We aimed to address the following questions: 1) Are Zoopagomycota truly rare in the environment?, 2) Are methodological biases (such as marker choice and fragment length) inhibiting the detection of Zoopagomycota?, and 3) Are both factors contributing to the absence of Zoopagomycota in metabarcoding studies? We hypothesized that: 1) fragment length amplification and sequencing bias would decrease the detection of Zoopagomycota species, 2) the LR22F/LR3 primer pair would outperform the LROR/LR3 primer pair in terms of OTU clustering and phylogenetic reconstruction, and 3) taxonomy assignment methods would strongly differ in the identification of Zoopagomycota and other EDF OTUs.

## Materials and Methods

*Zoopagomycota fungi in the mock community* − A mock community comprised of Zoopagomycota fungi was created by combining equilibrated aliquots of genomic DNA from species of Dimargaritales, Kickxellales, and Zoopagales. DNA of mock community members was obtained from cultures grown from the University of Florida Gerald Benny culture collection or received from collaborators (TABLE 1). Species of Dimargaritales, *Piptocephalis,* and *Syncephalis* are haustorial mycoparasites and were grown in dual cultures with their host fungi (Benny et al., 2016b). Accordingly, the DNA extracts from these cultures also contained an unknown quantity of host DNA. Similarly, genomic DNAs from the Davis et al. (2019) study were single cell genomes amplified by multiple displacement methods and also contained DNA from both the fungi and their host organisms. Species of Kickxellales are saprotrophic and were grown axenically (Benjamin, 1958). A preliminary test of ITS1F/ITS2 primers on DNA from *Piptocephalis* and*Syncephalis* species mixed with soil DNA indicated that increased ITS1 length in these taxa resulted in reduced amplification and sequencing (SUPP FIG 1). We included those taxa and additional taxa with known length variation in our mock community to further examine these potential biases across different markers. The final community contained 30 isolates of Zoopagomycota fungi (TABLE 1). We also generated reference Sanger sequences from individual mock community members using primer pair LROR/LR5 (Hopple and Vilgalys 1994; Vilgalys and Hester 1990) for LSU and primer pair ITS1F/ITS4 (Gardes & Bruns, 1993; White et al., 1990) for ITS. Reference sequences were verified by BLAST analysis against NCBI GenBank and with phylogenetic reconstruction (data not shown). A non-biological, equimolar DNA mock community which consisted of a mixture of 12 synthetic single-copy sequences (SYNMO) was included alongside the ITS1 samples to help detect index bleed between samples and evaluate bioinformatic parameters (Palmer et al., 2018). All OTUs recovered from the mock community samples were submitted for BLAST searches to assess the taxonomic identity of the OTUs for comparison against the bioinformatic output.

*Environmental samples* − We collected environmental samples from five sites in CA and two in FL (SUPP TABLE 1). At each site five substrates were collected: 1) bulk water from a freshwater stream or pond (water), 2) saturated sediment from the edge of the water body (mud), 3) the upper soil layers and leaf litter, consisting of the visible organic layer (topsoil), 4) the mineral soil layers below the topsoil (deep soil), and 5) microinvertebrates collected from the soil samples using Baermann funnels (invertebrates). At each site five replicates of each sample type were collected. The topsoil and deep soil replicates were collected at increasing distances from the water source along a 25 m transect (i.e. sample 1 was closest to the water and soil sample 5 was furthest from the water). For each sample approximately 15-25 mL of soil was collected into sterile 50 mL tubes and filled with sterile 2x Cetyl Trimethyl Ammonium Bromide lysis buffer (CTAB) to 30-35 mL. Water samples were collected in 950 mL sterilized Mason jars by dipping the jar into the water from the embankment. Water samples included some sediment and debris present on the bottom or floating on the top of the water. Large debris such as sticks, rocks, and clumps of leaves were removed. Vacuum filtration and a sterile Büchner funnel were used to filter the water through filter paper (6 µm pore size). After filtration, the filter papers were immediately placed in sterile 50 mL tubes filled with CTAB. No water samples were collected from the "Sweeney wash" site in California because there was no standing water although mud samples were collected from a wet depression between rocks. Microinvertebrates (protists,

nematodes, tardigrades, etc.) were collected using Baermann funnels. A 50/50 mixture of topsoil and deep soil was added to a funnel, filled with sterilized water, and covered with Parafilm. One mL of water was collected in a sterile tube after 24 hours incubation and stored in a -20 C freezer. A second mL of water was collected during the second 24-hour period. The two samples were centrifuged at 15000 x g for 15 minutes, the excess water was drained, and then the samples were combined into one tube with CTAB for DNA extraction. For the three "Sweeney" desert sites in California, three rather than five Baermann funnel replicates were obtained due to limited funnel availability. Five microinvertebrate replicates were performed for all other sites. Baermann funnel samples are enriched for potential hosts of Zoopagales parasites and were used to target these taxa.

*DNA Extraction* − DNA extraction followed a modified CTAB protocol (Gardes & Bruns, 1993). Topsoil, deep soil, and filter papers from the water samples were subjected to several cycles of freezing and thawing in 50 mL tubes with CTAB prior to DNA extraction. Several sterilized glass beads were then added, and samples were shaken for 1 minute at 1500 RPM in a 1600 MiniG tissue homogenizer (SPEX, Metuchen, NJ, USA). Two mL of CTAB was collected from each 50 mL tube and placed in sterile microcentrifuge tubes. Samples in CTAB were incubated with a 1:1 mixture of phenol:chloroform overnight and then washed with an additional chloroform step. The remainder of the CTAB protocol was performed without modification. Extractions from cultures followed the methods of Reynolds et al. (2019) and also used the CTAB protocol. Following extraction, DNA yield was estimated by a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Samples with genomic DNA concentrations >200 ng/ul were further cleaned with the DNeasy PowerClean Pro cleanup kit (Qiagen, Germantown, MD, USA); samples with low DNA yield (<200 ng/uL) were not cleaned due to loss of DNA during the clean-up process.

*Primer Selection* − A multiple sequence alignment containing Dikarya, Mucoromycota, and Zoopagomycota sequences from GenBank, as well as the newly sequenced Zoopagomycota species, was created in Mesquite (Maddison & Maddison, 2019) and aligned with MUSCLE (Edgar, 2004). This reference alignment was used to evaluate the number of mismatches between published primer sequences and Zoopagomycota fungi. After testing several modified primer combinations on four samples, we chose the primers ITS1F and ITS2 (White et al., 1990) for the ITS1 region for further comparisons. For LSU, we chose the LROR/LR3 (Hopple & Vilgalys, 1994) primer set, which has been successfully used in metagenomics studies by processing only the forward reads (Benucci et al., 2019; Bonito et al., 2014; Johansen et al., 2016). We also included a modified forward primer LR22F (5´-GAGACCGATAGHRHACAAG-3´) used in combination with reverse primer LR3. LR22F is the reverse compliment of primer LR22 to which we added three degenerate positions to maximize compatibility with EDF. This primer is similar to LR22R (Mueller et al., 2015, 2016) but shifted upstream 8 bp because target Zoopagomycota taxa in our alignment had mismatches to LR22R (SUPP FIG 2). Hereafter we refer to these primer sets by the forward primer: ITS1F, LROR or LR22F.

*Library Preparation* − We prepared the ITS1F Illumina library using the thermocycling protocols of Truong et al. (2019). Briefly, we used Phusion high fidelity polymerase (ThermoFisher Scientific, Waltham, MA, USA) and a dual-indexing approach with the following modifications. Amplification and sequencing with LROR were the same as for ITS1F except that the thermocycling program for the initial amplification step was 95°C for 1 minute, step down of -0.1°C per cycle from 55°C, and 72°C for 1:30 with a total of 30 cycles and a final elongation step of 72°C for 10 minutes. Temperature gradient tests were performed for LR22F to determine the best annealing temperature. The thermocycling program for LR22F was the same as for LROR except that the annealing temperature was a step down of -0.2°C per cycle starting from 65°C. All samples were amplified in three separate replicates and the replicates were combined prior to the indexing reaction. Negative PCR controls were included in all reactions. Amplicons from both the initial amplification and index attachment reaction were verified on 1.5% agarose gels. Indices were added using the Nextera XT index kit v2 (Illumina, San Diego, CA, USA) in a separate PCR step with the following protocol: GoTaq Green master mix (Promega, Madison, WI, USA) was used for the reaction and the thermocycling program included nine cycles at 94°C for 1 minute, 55°C for 1 minute, 72°C for 1:30, and a final elongation step of 72°C for 10 minutes. Indexed PCR products were pooled into groups of three based on similar band intensity and then cleaned with the Select-a-Size DNA Clean and Concentrator kit (Zymo, Irvine, CA, USA). All samples

were then quantified using a Qubit 4.0 Fluorometer (Invitrogen, Waltham, MA, USA), equilibrated, and combined for the final library. Libraries were further purified with the Agencourt AMPure XP kit (Beckman Coulter, Brea, CA, USA) to remove primer dimers and then sequenced with the Illumina MiSeq 300 bp PE protocol using V3 chemistry (Illumina, San Diego, CA, USA) at either the UF Interdisciplinary Center for Biotechnology Research or the UC Riverside Institute for Integrative Genome Biology. Raw data are available at NCBI's Sequence Read Archive (BioProject PRJNA660245).

Bioinformatics – All sequencing data were processed with the AMPtk pipeline v 1.4.1 (Palmer et al., 2018). Up to two nucleotide mismatches were allowed in each primer, a maximum of two expected errors were allowed during demultiplexing and quality filtering, the maximum length was set at 600 bp, a standard index bleed of 0.5% was set, and reference-based chimera filtering was used during clustering. The 600 bp maximum length refers to the truncation of reads for downstream processing in AMPtk. Reads are only truncated if they are above the length cutoff. We tested several different length cutoffs and found that 600 bp returned the best results for consistent contig formation. For the ITS1F data, a minimum length of 150 bp was used, the clustering threshold was 97% using the UNOISE3 (Edgar 2010) algorithm, and the built-in UNITE ITS database in AMPtk was used for taxonomic identification via the hybrid method. The ITS OTUs were also assigned taxonomy with the RDP classifier for comparisons across primer sets. The LSU data from both primer sets were processed with a minimum length of 225 bp and a clustering threshold of 98% using the UNOISE3 method. Because some fungal species have ITS1 sequences under 200 bp, a shorter minimum length was used for ITS1F than LSU. A custom LSU database was installed in AMPtk for taxonomy assignment (see below) because the built-in LSU database was outdated (RDP training set 8) and no other comprehensive LSU database compatible with AMPtk was available. The clustering thresholds for the LROR and LR22F datasets were determined by evaluating which mock community output best matched the true community composition after processing the data at different cutoffs (97, 98, and 99% for both primer sets) and comparing UNOISE3 against DADA2 (Callahan et al., 2016). For both LSU data sets, UNOISE3 was used instead of DADA2 because DADA2 doubled the number of OTUs in the output and did not significantly improve recovery of the mock community. Additionally, using UNOISE3 enabled direct comparisons between datasets generated with the three primer sets. Any remaining OTUs detected in the negative controls after filtering were deleted from the OTU tables.

To create the LSU database, the RDP (training set 11) (Liu et al., 2012; Wang et al., 2007) and SILVA (LSU Ref 132) (Quast et al., 2013) fungal LSU reference FASTA files were downloaded, concatenated, and the taxonomy strings reformatted for use in AMPtk. Reference sequences from Zoopagomycota in the mock community were added to the database along with >200 sequences of animals, fungi, plants, and protists from GenBank. We selected these additional sequences based on BLAST results of OTUs in the dataset that were initially unidentified or misidentified by the databases. Due to limitations in formatting, the database is only compatible with the UTAX or DADA2 taxonomy assignment methods in AMPtk. The final updated database consisted of 115,382 dereplicated sequences and is freely available on OSF (https://osf.io/cz3mh/). We refer to this modified database as RDP+SILVA, and the program used for taxonomy assignment was UTAX. All three OTU datasets were also assigned taxonomy with the RDP classifier v 2.12 (Wang et al., 2007) executed in AMPtk using the ITS UNITE and LSU training sets, a commonly used methodology for metabarcoding studies.

Statistical analyses – Once OTU tables were obtained, further analyses were conducted in R (R Core Team, 2019) with scripts available at OSF (https://osf.io/cz3mh/). To check sampling coverage, rarefaction plots were made using the package iNEXT (Chao et al., 2014; Hsieh et al., 2020). Community comparisons were performed on the subset of OTUs identified as EDF and compared to all fungi. To visualize community differences among primer sets, the OTU table was converted into presence/absence format and $B$ sim dissimilarity matrices were calculated with the betadiver function in the vegan package (Oksanen et al., 2019) using the "w" method (Koleff et al., 2003). Non-metric multidimensional scaling (NMDS) ordinations were performed with the metaMDS function in vegan and plotted using ggplot2 (Wickham, 2016). These ordinations were visualized for the entire fungal community and on subsets of the EDF OTUs. NMDS ordinations were performed for each primer set separately and the scores were extracted as dataframe objects which

were then overlayed onto a single plot without modification of the coordinates. To compare the effect of the sampling location and sample type on fungal and EDF community composition, the dissimilarity matrices were analyzed with permutational multivariate analysis of variance (PERMANOVA - Anderson, 2001) of the *B* sim distance matrices (Anderson et al., 2006) using the adonis2 function in the vegan package. Because PERMANOVA tests are sensitive to dispersion of the samples within groups, an ANOVA analysis of the *B* sim distance matrices was conducted to test for significant differences between groups of interest. Significant ANOVA results of *B* sim output indicate that dispersion between groups may confound results from PERMANOVA analyses. Centroid differences were used for *B* sim (type = "centroid"), and both *B* sim and PERMANOVA used 9,999 permutations. To further examine similarities between primer sets, Mantel tests of correlation between the dissimilarity matrices of each primer set were conducted using Spearman's rank correlation and 999 permutations. Finally, taxonomy bar plots and alpha diversity metrics for all fungi and the subset of EDF were conducted in phyloseq (McMurdie & Holmes, 2013) or the R base functions. Alpha diversity measures were evaluated for fungal community differences based on site and sample type with ANOVA and Tukey's Honest Significant Differences using the agricolae package (de Mendiburu, 2020). Mantel tests of fungal OTUs utilized a subset of the data that contained only the 127 samples (118 for EDF) that worked for all three primer sets.

We selected 50 OTUs from the LROR dataset and 50 OTUs from the LR22F dataset to study using phylogenetic analyses. We selected those that had the greatest number of reads, were detected in more than one sample, and were identified only as kingdom Fungi with our pipeline (i.e. the UTAX algorithm and RDP+SILVA database). The OTUs were also subjected to BLAST searches and OTUs with high matches to non-fungal sequences were not included. These 100 unknown OTUs were added to a sequence alignment of Zoopagomycota, other EDF taxa, and additional Dikarya. As a check on the taxonomic identification by our pipeline, OTUs classified within the Zoopagomycota (Kickxellomycotina or Zoopagomycotina) were included in a smaller alignment and processed the same way as the larger alignment. The sequences were aligned with MUSCLE, ambiguously aligned regions were excluded, and Maximum Likelihood analyses were performed in RAxML v 8 (Stamatakis, 2014) using the GTRGAMMA model and 1,000 bootstraps. The unknown OTUs were also examined using BLAST searches against GenBank using both default parameters and excluding uncultured/environmental sample sequences and the results were compared to the placement of the OTUs in the phylogeny. Resulting figures were modified in FigTree v 1.4.3 (Rambaut, 2012) and InkScape v 0.92.2 (https://inkscape.org/en/).

## RESULTS

*Comparisons of sequencing and bioinformatic processing*– The ITS1F dataset included 146 samples (105 CA, 42 FL), had the most reads (14.2 million), and resulted in 7,609 OTUs. The LR22F dataset included 142 samples (106 CA, 36 FL), had the fewest reads (7.4 million), and resulted in 10,028 OTUs. The LROR dataset included 144 samples (107 CA, 37 FL), had an intermediate number of reads (9.1 million) and resulted in 5,786 OTUs (FIG 1). The ITS1F dataset had the highest percentage of reads discarded due to primer incompatibility (6.37%) while the LR22F set had the lowest percentage of primer incompatibility (0.38%), but the highest percentage of reads discarded due to short length (5.76%). Data from the LR22F primer set were mostly assembled into contigs that used both forward and reverse reads (FIG 1). In contrast, pairing between forward and reverse reads for ITS1F data was widely variable, whereas reads from the LROR primer set generally could not be compiled into contigs. This resulted in the majority of LROR OTUs being comprised of only forward reads (283 bp). Rarefaction plots indicate sufficient sampling from each state for each primer set (SUPP FIG 3).

The proportion of fungal OTUs was different for each of the three primer pairs and the two taxonomy assignment methods (FIG 2, SUPP FIG 4). The ITS1F dataset recovered the most fungal OTUs followed by LROR and LR22F according to the hybrid (for ITS1F) and UTAX (for LSU) methods of taxonomy assignment. The RDP classifier assigned 100% of ITS1F OTUs to Fungi but assigned a greater proportion of LROR and LR22F OTUs to non-fungal groups than the UTAX method (47.2-54% RDP vs 38.8-33.2% UTAX) (FIG 2 C). The majority of fungal reads from all three datasets were assigned to Ascomycota followed

by Basidiomycota but the three datasets differed in the proportion of sequences that were assigned to EDF groups (FIGS 2, 4). The maximum OTU length was similar for all primer sets (ITS1F 534 bp, LR22F 548 bp, LROR 546 bp). Correlation plots of OTU length versus read number indicate that these variables are significantly negatively related for ITS1F, but not for the LSU datasets (SUPP FIG 5). Finally, the ITS1F dataset had the lowest percentage of unidentified OTUs with 1.9%, followed by LROR with 23.8%, and LR22F with 35.5% using the hybrid and UTAX methods, respectively. The RDP classifier recovered fewer unidentified OTUs with 0% for ITS1F, 18.3% for LR22F, and 24.1% for LROR.

Within Fungi, several orders of Dikarya were dominant across all primer sets and samples (SUPP FIG 6; SUPP TABLE 3). The greatest differences between primer sets were among less OTU-rich groups that were recovered by only one or two of the three primer sets (FIG 3; SUPP FIG 6). The LROR and LR22F datasets recovered more EDF OTUs from Blastocladiales, Calcarisporiellales, Chytridiales, Cladochytriales, Endogonales, Entomophthorales, Gromochytriales, Monoblepharidales, Neocallimastigales, Zoopagales, and Microsporidia than ITS1F (FIG 3). The LROR and LR22F datasets also recovered more OTUs from fungal-like organisms in class Oomycetes (kingdom SAR) and slime molds in Physariida (kingdom Amoebozoa). In other cases, the ITS1F dataset had several times the number of OTUs than either of the LSU datasets. For example, Rozellomycota was the dominant taxon among EDF for ITS1F, but Chytridiomycota was dominant for both LSU markers (FIG 3). Likewise, the ITS1F dataset had 36 OTUs assigned to Kickxellales compared to five or fewer for each of the LSU primer sets. This mirrors the inflation of Kickxellales in the mock community in the ITS1F dataset (see below).

*Comparison of mock communities* − Inspection of the ITS1F mock community shows some likely errors in the OTU taxonomic assignment (TABLE 1). The AMPtk hybrid taxonomy assignment method identified two ITS1F OTUs as Stramenopiles (SAR), but each had BLAST matches to *Acaulopage* (Zoopagales), a mock community member (73% coverage, 91.5% identity and 72% coverage, 78.08% identity). The OTU identified as *Dimargaris cristilligena* had a BLAST match of 100% coverage and 97.4% identity to *Cokeromyces recurvatus,*the host of this mycoparasite. Five other OTUs found in the ITS1F Zoopagomycota mock community were identified as Basidiomycota, Metazoa, or Haptista. These Dikarya and non-fungal OTUs were not found in the negative controls and they were not returned in the mock communities sequenced with the LSU primers. These OTUs could have been amplified from the mixed genomic DNA present in the non-axenic mock isolates. The ITS1F hybrid dataset also returned 12 OTUs assigned to Kickxellales in the mock community whereas only nine taxa were actually included. Similarly, the ITS1F dataset had only three Zoopagales OTUs even though 21 isolates were originally added. The RDP classifier method identified 5 putative Mucoromycota host fungi OTUs and one Kickxellales OTU from the mock community. The remaining 21 OTUs were classified only to kingdom Fungi by RDP.

Mock community recovery was more accurate with the LSU datasets using the UTAX taxonomy assignment and RDP+SILVA database than the ITS1F dataset (TABLE 1). Both LSU primer sets recovered almost all the members of the mock community except that the LR22F dataset identified one less *Coemansia* OTU. Both LSU primer sets also recovered OTUs assigned to putative Mucoromycota host fungi of the mycoparasites included in the mock. Comparison of UTAX against the RDP+SILVA LSU database to the RDP classifier shows that few of the mock community members were identified by RDP (TABLE 1). Many mock isolates were classified as Metazoa or remained unclassified with RDP but were accurately identified as fungi by UTAX and the RDP+SILVA LSU database. Across all primer sets, the Zoopagales mock members were underrepresented with several isolates remaining undetected by all three primer sets. There are multiple possible reasons these taxa did not amplify or sequence well: 1) we found that *Acaulopage dichotoma* , *A. tetraceros* , and*Stylopage* species lack part of the ITS2 priming site (see alignment files at https://osf.io/cz3mh/), 2) amplification competition from shorter host DNA fragments present in the community, and/or 3) these fungi may have uncharacterized sequence features that reduce amplification, such as high G:C content (Dutton et al., 1993).

Ordination plots and statistical analyses – Community analyses are based on the subset of OTUs identified as Fungi or the subset of OTUs identified as EDF. The ITS1F primer set recovered 6,140 fungal OTUs, 1,757

of which were EDF, LROR recovered 2,095 fungi (369 EDF), and LR22F recovered 3,126 fungi (447 EDF). There were no observable differences in the fungal community recovered between primer sets based on the grouping of samples in the NMDS plots (FIG 4) (individual plots for each primer set separately are given in SUPP FIG 7). However, fungal communities from CA and FL were distinct with additional partitioning based on sampling site. The Evey Canyon and San Jacinto sites clustered separately from the Mojave Desert Sweeney sites in CA, whereas the two FL sites overlapped (FIG 4B, D). For the EDF, CA site clusters were separated in the ITS1F dataset, but those communities overlapped in the LSU datasets (FIG 4 A). PERMANOVA analyses found significant effects of site and sample type on the fungal communities among each of the primer sets (TABLE 2), but the effect size ($R^2$) was small. Generally, sample type had a greater effect on EDF and fungal communities from FL, whereas site had a stronger impact on communities in CA. The ANOVA results of the$B$ sim matrices for CA were all significant except for ITS1F site (fungi) and sample (EDF). Mantel test comparisons between primer sets were all significant for both fungi and EDF, indicating that the recovered communities were significantly correlated between the three datasets (TABLE 3). The correlation (R) was less than 0.50 for fungal ITS1F vs. LROR and LROR vs. LR22F and less than 0.60 for all EDF comparisons. The highest Mantel R statistic was for fungal ITS1F vs. LR22F, with a value of 0.71.

Plots of fungal richness between each primer set, site, and sample type showed that overall fungal richness was similar across all three primer sets, but soil samples generally had the highest diversity (SUPP FIG 8). Evey Canyon and San Jacinto had among the highest OTU richness for all sample types across all three primers. Fungal richness was lower for invertebrate samples across all primer sets but was higher for ITS1F than either LSU primer set. However, EDF diversity patterns differed. For ITS1F, many water, mud, and invertebrate samples had greater EDF diversity than soil (FIG 5). Conversely, soil and mud from Sweeney sites in the Mojave Desert had among the highest EDF diversity for both LSU markers (FIG 5). Tukey's test results varied for EDF and all fungi and by primer (SUPP TABLE 4) and separation of EDF communities by sample type and site are also observable in the NMDS plots (SUPP FIG 7).

*Phylogenetic reconstruction of LSU OTUs* – The 50 LROR and 50 LR22F OTUs identified only as "Fungi" were added to the sequence alignment that included 436 taxa. After exclusion of ambiguous sites, the LSU alignment contained 1,104 characters and OTUs had a final length of 216-239 bp for LR22F and 186-230 bp for LROR. Figure 7 shows the phylogeny and SUPP TABLE 5 lists the LSU OTUs used in the phylogeny and BLAST results for each OTU. Backbone nodes of the phylogeny were mostly unsupported whereas nodes near the tips had higher bootstrap support (>70). Both primer sets recovered monophyletic clades of OTUs that had BLAST matches to protist sequences (SUPP TABLE 5), but the LR22F dataset had fewer than LROR (FIG 6). The majority of these LROR OTUs had higher BLAST identity scores to the protist sequences than the LR22F OTUs, a maximum of 81% coverage and 100% identity for LROR, but only 20% coverage and 98.86% identity for LR22F. The LROR OTUs that had BLAST matches to protists were resolved in two different clades, one with three OTUs that had matches to Stramenopiles, and a larger clade with matches to Rhizaria. The LR22F protist clade was nested within the Chytridiomycetes, but all the matches (except the one mentioned above) had identity scores in the 75-78% range. Most fungal OTUs from both primer sets were resolved in the Chytridiomycetes and Orbiliomycetes. The OTUs placed in the Orbiliomycetes had close BLAST matches to Orbiliomycetes. In contrast, many OTUs placed in the Chytridiomycetes had matches to other fungal orders. Only LROR OTUs were placed within Aphelidiomycetes, and clades in the Basidiomycota, Eurotiomyctes, Umbelopsidomycetes, and Zoopagomycetes. The Archaeorhizomycetes, Endogonomycetes 1, and Sordariomycetes contained LR22F OTUs but none from LROR. The Glomeromycetes contained four LROR OTUs and one LR22F OTU. In the Zoopagomycota-only phylogeny, the two subphyla are recovered as polyphyletic, contrary to other studies (Davis et al., 2019), with the Dimargaritales and *Ramicandelaber* nested within the Zoopagomycotina (FIG 7). Additionally, the Harpellales are sister to the two subphyla rather than nested within Kickxellomycotina as found by other studies (Wang et al., 2019). Only OTU 6654 (LR22F) did not place in a clade matching its taxonomic classification from the taxonomy pipeline.

## DISCUSSION

Taxonomic assignment of OTUs – Recovery of EDF communities is impacted more by methodological choices during metabarcoding sampling than Dikarya communities. In particular, target fragment length and taxonomy assignment method can each have a profound impact on EDF detection. We found considerable differences in taxonomic identifications between taxonomy methods for all primer sets. This was even true at the Kingdom level where LSU OTUs were more likely to be identified as non-fungal by the RDP classifier (47-54%) compared to our RDP+SILVA database (33-38%). The ITS1F dataset had the greatest number of OTUs identified as Fungi. This result was expected due to differences in primer specificity (i.e. the ITS1F/ITS2 primer combination is more fungal-specific but has known biases for Dikarya – Bellemain et al., 2010; Tedersoo et al., 2015) and the completeness of the ITS versus LSU databases. Because the ITS databases have >1,000,000 reference sequences compared to <200,000 for LSU databases, we expected more accurate taxonomic assignment for ITS1F OTUs.

In the mock community analyses, both LSU datasets more closely recapitulated the community than ITS1F (TABLE 1). Errors in the taxonomy assignment of ITS1F OTUs from the mock community indicate misidentification of reference sequences (e.g. *Dimargaris* ) and also a lack of reference sequences for some taxa (e.g. Zoopagales) (TABLE 1). Correlation plots of fungal OTU length and taxonomic identity score had significant negative relationships for the LSU primer sets (SUPP FIG 9), but a positive relationship for ITS1F. The negative relationship for LSU likely reflects the greater representation of the shorter length reference sequences from Dikarya species in the databases than longer non-Dikarya. However, while there was a negative relationship between OTU length and read number for ITS1F, there was no significant correlation for the LSU datasets (SUPP FIG 5). This implies that taxonomic representation in the reference databases rather than OTU length has a greater impact on taxonomy scores of LSU OTUs. However, our RDP+SILVA LSU database substantially improved identification of Zoopagomycota isolates in the mock community compared to the RDP classifier (TABLE 1) with ≤10% of mock members identified with RDP versus 63% with RDP+SILVA. In contrast, ITS1F recovered 27% of the mock members using the hybrid taxonomy method. It is important to note, however, that populating our RDP+SILVA database with additional non-fungal sequences from GenBank improved the accuracy of determining fungal versus non-fungal sequences compared to RDP. The supported placement of Zoopagomycota OTUs in our reduced phylogeny also reinforces the accuracy of the taxonomic classifications made by our pipeline (FIG 7). Therefore, a robust reference database should include a diversity of eukaryotic sequences, especially for LSU because primers for this region are often less fungal-specific.

These results demonstrate the profound effect that reference databases have on the classification of OTUs. Other taxonomic assignment approaches, like BLAST followed by MEGAN (Huson et al., 2011) or the Statistical Assignment Package (Munch et al., 2008), have the potential to improve LSU OTU identification (Porter & Golding, 2012). However, many of our unidentified OTUs had best matches to other unidentified sequences in GenBank (e.g. SUPP TABLE 5) or best matches that did not reflect the phylogenetic placement of the OTU, indicating that BLAST and reference-based methods cannot completely alleviate identification problems (Lücking et al., 2020). The disparity between named fungal species and unnamed environmental sequences is substantial and available data suggest that many unidentified sequences represent EDF, including Zoopagomycota (Lazarus & James, 2015; Tedersoo et al., 2017, 2020). Even within a relatively small order, Zoopagales, 13 out of 22 genera lack any sequence data (Davis et al., 2019b). Although much has been done to improve fungal LSU databases (e.g. Vu et al., 2019; Hanafy et al., 2020), further additions and curation can bring LSU on par with ITS as an rDNA metabarcoding marker. Until intensive efforts are made to curate and fill the taxonomic gaps within databases, it is clear that taxonomic assignment issues will be problematic irrespective of improvements in sequencing and bioinformatics. For example, the long reads and simultaneous sequencing of multiple rDNA markers at once offered by PacBio technology was not able to entirely overcome the pitfalls of reference-based taxonomic assignment (Furneaux et al., 2021; Heeger et al., 2018).

Primer comparison – Comparisons between metabarcoding datasets from different studies are challenging due to variable methods of sample collection, PCR amplification, sequencing, and bioinformatics. However, our results using data from >127 environmental samples support the broad pattern of fungal community

congruence between ITS and LSU markers as evidenced by significant Mantel tests (TABLE 3) and found by others (Amend et al., 2010; Benucci et al., 2019; Bonito et al., 2014; Brown et al., 2014; Johansen et al., 2016; Mota-Gutierrez et al., 2019; Nelson & Shaw, 2019; Skelton et al., 2019; Xue et al., 2019). Our NMDS plots (FIG 4; SUPP FIG 7) demonstrate that the ITS1F, LROR, and LR22F datasets recovered similar fungal communities and detected mostly the same OTU-rich lineages (SUPP FIG 6). However, there were discrepancies among rarer groups, like some chytrid orders for which more LSU OTUs were detected than ITS1F (FIG 3). Similarly, Nelson and Shaw (2019) and Benucci et al., (2018) reported greater numbers of Chytridiomycota OTUs with LROR than ITS. Conversely, the ITS1F marker inflated the number of OTUs assigned to the Kickxellales in the mock community; we originally added nine taxa but recovered 12 OTUs. This could be a result of biological (e.g. intraspecific rDNA copy number variation) and/or methodological (e.g. inappropriate OTU clustering identity threshold) factors. Artificial inflation of the number of ITS OTUs has been shown for various taxa in mock communities in metabarcoding studies (Castaño et al., 2020; De Filippis et al., 2017; Jusino et al., 2019; Nguyen et al., 2015; Větrovský et al., 2016). These results underscore the importance of mock communities for detecting methodological errors and refining bioinformatic parameters such as clustering thresholds (Caporaso et al., 2011; Palmer et al., 2018; Taylor et al., 2016). Furthermore, taxonomic assignments of ITS OTUs cannot be tested with phylogenetic analyses. Thus, groups with dramatic differences in representation between markers (e.g. Rozellomycota in our dataset, FIG 3) cannot easily be evaluated for accuracy.

We further aimed to test LSU primer pairs for their ability to amplify fungi broadly, and Zoopagomycota fungi specifically, as well as compare their performance in the bioinformatics pipeline. The LR22F dataset had more fungal and total OTUs than LROR and fungal diversity analyses found significant differences between groups for LR22F not found with LROR (SUPP TABLE 4; SUPP FIG 8). Similar results were found by Mueller et al. (2016) for the related LR22R primer, which recovered richness estimates closer to ITS than LROR. We found that the forward and reverse reads from the LR22F dataset were consistently paired into contigs, contrary to both the ITS1F and LROR primer sets (FIG 1). The LROR target fragments for the mock community members were commonly >700 bp (TABLE 1), well beyond the sequenceable length of Illumina MiSeq chemistry. The resulting data are therefore almost entirely restricted to the forward reads, resulting in significant data loss. However, paired reads have several advantages over unpaired reads: more data are utilized, overlapping sequences reduce sequence errors, and longer sequences reduce problems during OTU clustering and taxonomy assignment (Bartram et al., 2011; Truong et al., 2019). Furthermore, longer sequences are more accurately identified using bioinformatic methods (Liu et al., 2012; Porras-Alfaro et al., 2014; Porter & Golding, 2012). As a result, the taxonomic identification of LROR OTUs may be less reliable than the longer LR22F OTUs. We also found that sequence length influenced our phylogenetic analysis where the longer LR22F OTUs were generally placed with higher resolution than shorter LROR OTUs (FIG 6). Twenty-two of the 50 LR22F OTUs were placed in a clade that matched the fungal class of their BLAST match, compared to only 15 of the 50 LROR OTUs. Conversely, both LROR and LR22F OTUs resolved well in the smaller Zoopagomycota tree (FIG 7), and the majority were placed in clades that correspond to the taxonomy assignment output from the UTAX/RDP+SILVA pipeline. These results illustrate the utility of phylogenetic reconstruction of LSU OTUs for identifying potential divergent EDF fungal sequences as well as sequence artifacts or taxonomic errors that need further investigation (Glass et al., 2013). For example, the OTUs that were placed within the Orbiliomycetes (and had high BLAST matches to Orbiliomycetes) indicate that the corresponding reference sequences in the database could be identified beyond kingdom to increase the accuracy of the taxonomy.

*Methodological biases impacting detection of EDF: the example of Zoopagomycota* – We found additional evidence that target region length (for both ITS1 and LSU fragments) strongly affects the metabarcoding process at different steps. For ITS1F, the strongest bias putatively occurs during PCR when shorter fragments are preferentially amplified (Castaño et al., 2020; Jusino et al., 2019; Palmer et al., 2018). In our initial experiments with the ITS1F primer set, Zoopagomycota species with the longest ITS1 ([?]400 bp) were not recovered from mixed samples. This was true despite adding DNA from those species at twice the concentration of the "background" DNA (SUPP FIG 1). This pattern was reiterated in mock community

results (TABLE 1) where target species with the longest ITS1 ([?]400 bp) were not detected. This bias towards short fragments is also taxonomically biased. Many species of Ascomycota, Basidiomycota, and Mucoromycota have short ITS1 regions (<200-300 bp) (Bellemain et al., 2010; Bokulich & Mills, 2013), with a few notable exceptions (e.g. Cantharellales may have ITS1 >1,000 bp – Feibelman et al., 1994). On the other hand, most Zoopagomycota species we examined have an ITS1 >300 bp with significant variation between taxa. For example, *Piptocephalis* and *Syncephalis* species have ITS1 that ranges 300 – 800 bp (Lazarus et al., 2017; Reynolds et al., 2019) and we have also found variation among *Coemansia* species (TABLE 1). However, the *Coemansia* isolates had shorter ITS1 than many other mock members, which likely contributed to their overrepresentation in the ITS1F results. Species of Harpellales have extreme variation with an ITS1 range of 250–1,000 bp (Gottlieb & Lichtwardt, 2001). Similar length variation occurs in the ITS2 region and in some cases ITS2 is longer than ITS1, such as in some Harpellales that have 1,100 bp for ITS2 but only 500 bp for ITS1 (Gottlieb & Lichtwardt, 2001). Although the fragment length for LR22F is variable among Zoopagomycota, the variation is lower than for ITS1F. In our mock community, the difference in fragment length between the longest and the shortest*Syncephalis* species was only 45 bp for LR22F versus 508 bp for ITS1F. Likewise, the range among *Piptocephalis* species was 85 bp for LR22F versus 247 bp for ITS1F.

Beyond sequencing and bioinformatics, the biology of Zoopagomycota must also be considered. The symbiotic nature of Zoopagomycota fungi means that their abundance in any given sample is linked with the abundance of their host organisms and dependent on host/parasite interactions. As a result, the distribution of these fungi is likely patchy and highly variable through time, lowering the probability of their detection from any single sample. Little is known about the host/parasite dynamics among Zoopagomycota parasites, making the choice of locations and sample types for metabarcoding less straightforward. For example, although the mycoparasitic species can be isolated from soil, they also are frequently isolated from dung. Are these species mainly coprophilic, or do they actively attack hosts in the soil environment as well? How long do their spores persist in the soil? Similarly, Zoopagales species that attack microinvertebrates have been isolated from wet substrates like moss and decaying plant material (Drechsler 1938; Duddington 1955). We attempted to concentrate potential hosts of these fungi and thereby increase our chance of detecting them by using Baermann funnels. However, these invertebrate samples were generally dominated by Chytridiomycota OTUs followed by Blastocladiomycota and Mucoromycota (FIG 3). A small number of Zoopagales OTUs were detected from other sample types by each primer set, and phylogenetic analyses supported their identification (FIG 7). Nonetheless, the number of OTUs detected is still less than the number of isolates recovered from these sites using specialized culturing techniques. For example, Benny et al., (2016) found that 46% of 520 soil samples (mostly from Florida) contained at least one *Syncephalis* species, but we only recovered one *Syncephalis* OTU from two samples using the LROR primer set. Likewise, the spores of Harpellales fungi are thought to pass between hosts through transmission in the water column (Lichtwardt 1986). However, we were unable to detect Harpellales OTUs from water or mud samples, which were mostly dominated by chytrid OTUs. Species of Kickxellales grow axenically and can be isolated from soil or dung, but their trophic modes remain unclear. Although Kickxellales fungi are assumed to be saprotrophic, there are reports of some species growing on other fungi or in association with arthropods and many species exhibit fastidious growth in culture (Jackson & Deardon, 1948; Linder 1943; Kurihara et al., 2001, 2008). Furthermore, most species are assumed to be rare, but the dearth of reports could be an artifact of undersampling. Our results demonstrate that Kickxellales can be detected from soil (FIG 3) and phylogenetic analyses indicate the clade may be more diverse and more widely distributed than currently recognized (FIG 7).

The combined effects of methodological biases and environmental sample heterogeneity (with symbiotic Zoopagomycota likely having lower abundance) may have a synergistic impact on metabarcoding outcomes, leading to artificially inflated representation of some groups and absence of others. These biases are rooted in methodology and can affect any group of organisms with highly variable target fragment lengths and/or poor reference sequence representation. Castaño et al. (2020) found such a pattern in mock communities where longer fragments added in unequal proportions to shorter fragments were severely underrepresented

in Illumina MiSeq results while shorter fragments could be highly overrepresented (up to 57% greater output than input). Although PacBio RS II sequencing produced less severe discrepancies, a length bias was still detected (Castaño et al., 2020), an effect that can also be influenced by sample loading method (Tedersoo et al., 2017b). Our results demonstrate that both PCR length bias and lack of reference sequences severely impact the detection of Zoopagomycota from mixed samples. Other lineages of EDF are similarly involved in symbiotic associations and lack reference sequence data (e.g. Blastocladiomycota, Rozellomycota), indicating they are likewise often missed or misidentified in metabarcoding studies. Many unanswered questions about Zoopagomycota remain, such as their roles in microbial food webs and their effect on host populations. Metabarcoding has the potential to help unravel some of these mysteries but novel approaches are needed to overcome methodological biases, such as PCR-free on-array hybrid capture (e.g. Mamanova et al., 2010). Combined with targeted culturing approaches, improved environmental sampling methods can help illuminate the diversity and ecological roles of "dark matter fungi" (Grossart et al., 2016).

**Acknowledgements**

This work was supported by the National Science Foundation (DEB 1441677 to MES; DEB 1441715 to JES). Additional funding was provided by the Department of Plant Pathology, the USDA-NIFA under award number FLA-PLP-005289 and the Institute for Food and Agricultural Sciences at the University of Florida. We thank Tim James (University of Michigan) and William J. Davis (USDA Mycology and Nematology Genetic Diversity and Biology Lab) for sharing Zoopagales DNA samples used to generate reference rDNA sequences and in the mock community. We also appreciate the guidance and input from Gerald Benny (University of Florida) who provided access to the zygomycete culture collection and helped with the culture work. The environmental sampling was performed at the University of California Natural Reserve System James San Jacinto Mountains Reserve (doi: 10.21973/N3KQ0T) and Sweeney Granite Mountains Desert Research Center (doi: 10.21973/N3S942) and at the University of Florida Ordway-Swisher Biological Station (https://ordway-swisher.ufl.edu/) in Putnam County, Florida.

**Author Contributions**

NKR and MES planned the research. MES and JES provided funding for the research. NKR collected samples from California and Florida and JES helped arrange access to reserves and collect samples from California. NKR performed lab work with the assistance of MAJ. NKR and MAJ conceived the bioinformatics and analyses and NKR analyzed the data with troubleshooting help from MAJ. NKR wrote the manuscript which was thoroughly edited by MES and included input from all co-authors.

**Data Availability Statement**

Jupyter notebooks used for raw data processing in AMPtk, R scripts used for data analyses, and the RDP+SILVA LSU database FASTA file are available for download from OSF (https://osf.io/cz3mh/). The GenBank accession numbers for the Sanger sequences reported in this paper are available in Table 1. Illumina raw sequences are available on NCBI's sequence read archive (SRA) at BioProject PRJNA660245 for the main dataset and PRJNA428770 for the supplemental data.

**References**

Abarenkov, K, Zirk, A., Piirmann, T., Pöhönen, R., Ivanov, F., Nilsson, R., & Kõljalg, U. (2020). *Full UNITE+INSD dataset for Fungi* . UNITE Community. https://doi.org/https://doi.org/10.15156/BIO/786372

Abarenkov, Kessy, 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* (2), 517–525. https://doi.org/10.1111/nph.15301

Amend, A. S., Seifert, K. A., Samson, R., & Bruns, T. D. (2010). Indoor fungal composition is geographically patterned and more diverse in temperate zones than in the tropics. *Proceedings of the National Academy of Sciences of the United States of America* , *107* (31), 13748–13753. https://doi.org/10.1073/pnas.1000454107

Anderson, M. J., Ellingsen, K. E., & McArdle, B. H. (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters* , *9* (6), 683–693. https://doi.org/10.1111/j.1461-0248.2006.00926.x

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

Asemaninejad, A., Weerasuriya, N., Gloor, G. B., Lindo, Z., & Thorn, G. (2016). New primers for discovering fungal diversity using nuclear large ribosomal DNA. *PLoS ONE* , *11* (7), 1–15. https://doi.org/10.1371/journal.pone.0159043

Badotti, F., De Oliveira, F. S., Garcia, C. F., Vaz, A. B. M., Fonseca, P. L. C., Nahum, L. A., Oliveira, G., & Góes-Neto, A. (2017). Effectiveness of ITS and sub-regions as DNA barcode markers for the identification of Basidiomycota (Fungi). *BMC Microbiology* ,*17* (1), 1–12. https://doi.org/10.1186/s12866-017-0958-x

Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G., & Neufeld, J. D. (2011). Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Applied and Environmental Microbiology* , *77* (11), 3846–3852. https://doi.org/10.1128/AEM.02772-10

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 Microbiology* , *10* (189). https://doi.org/10.1016/j.enzmictec.2010.03.010

Benjamin, R. K. (1958). Sexuality in the Kickxellaceae. *Aliso* ,*4* (1), 149–169. https://doi.org/10.5642/aliso.19580401.14

Benjamin, R. K. (1959). The Merosporangiferous Mucorales. *Aliso, 4* (2), 321–433.

Benjamin, R. K. (1961). Addenda to "The Merosporangiferous Mucorales".*Aliso, 5* (1), 11-19.

Benny, G. L., Ho, H. M., Lazarus, K. L., & Smith, M. E. (2016). Five new species of the obligate mycoparasite *Syncephalis* (Zoopagales, Zoopagomycotina) from soil. *Mycologia* , *108* (6), 1114–1129. https://doi.org/10.3852/15-073

Benny, G. L., Smith, M. E., Kirk, P. M., Tretter, E. D., & White, M. M. (2016). Challenges and Future Perspectives in the Systematics of Kickxellomycotina, Mortierellomycotina, Mucoromycotina, and Zoopagomycotina. In D. Li (Ed.), *Biology of Microfungi* (pp. 65–126). Springer International Publishing. https://doi.org/10.1007/978-3-319-29137-6

Benucci, G. M. N., Bonito, V., & Bonito, G. (2019). Fungal, Bacterial, and Archaeal Diversity in Soils Beneath Native and Introduced Plants in Fiji, South Pacific. *Microbial Ecology* , *78* (1), 136–146. https://doi.org/10.1007/s00248-018-1266-1

Bokulich, N. A., & Mills, D. A. (2013). Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Applied and Environmental Microbiology* , *79* (8), 2519–2526. https://doi.org/10.1128/AEM.03870-12

Bonito, G., Reynolds, H., Robeson, M. S., Nelson, J., Hodkinson, B. P., Tuskan, G., Schadt, C. W., & Vilgalys, R. (2014). Plant host and soil origin influence fungal and bacterial assemblages in the roots of woody plants. *Molecular Ecology* , *23* (13), 3356–3370. https://doi.org/10.1111/mec.12821

Brown, S. P., Rigdon-Huss, A. R., & Jumpponen, A. (2014). Analyses of ITS and LSU gene regions provide congruent results on fungal community responses. *Fungal Ecology* , *9* (1), 65–68. https://doi.org/10.1016/j.funeco.2014.02.002

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. *Nature Methods* , *13* (7), 581–583. https://doi.org/10.1038/nmeth.3869

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America* , *108* (SUPPL. 1), 4516–4522. https://doi.org/10.1073/pnas.1000080107

Castaño, C., Berlin, A., Brandström Durling, M., Ihrmark, K., Lindahl, B. D., Stenlid, J., Clemmensen, K. E., & Olson, Å. (2020). Optimized metabarcoding with Pacific biosciences enables semi-quantitative analysis of fungal communities. *New Phytologist* . https://doi.org/10.1111/nph.16731

Chao A, Gotelli NJ, Hsieh TC, Sande EL, Ma KH, Colwell RK, Ellison AM (2014). "Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies." Ecological Monographs, 84, 45–67.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R., & Tiedje, J. M. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research* , *42* (D1), 633–642. https://doi.org/10.1093/nar/gkt1244

Corsaro, D., Köhsler, M., Wylezich, C., Venditti, D., Walochnik, J., Michel, R. (2017). New insights from molecular phylogenetics of amoebophagous fungi (Zoopagomycota, Zoopagales). Parasitology Research, 117, 157-167. https://doi.org/10.1007/s00436-017-5685-6

Corsaro, D., Walochnik, J., Venditti, D., Steinmann, J., Müller, K. D., & Michel, R. (2014). Microsporidia-like parasites of amoebae belong to the early fungal lineage Rozellomycota. *Parasitology Research* ,*113* (5), 1909–1918. https://doi.org/10.1007/s00436-014-3838-4

Davis, W. J., Amses, K. R., Benny, G. L., Carter-House, D., Chang, Y., Grigoriev, I., Smith, M. E., Spatafora, J. W., Stajich, J. E., & James, T. Y. (2019). Genome-scale phylogenetics reveals a monophyletic Zoopagales (Zoopagomycota, Fungi). *Molecular Phylogenetics and Evolution* , *133* , 152–163. https://doi.org/10.1016/j.ympev.2019.01.006

Davis, W. J., Amses, K. R., James, E. S., & James, T. Y. (2019b). A new 18S rRNA phylogeny of uncultured predacious fungi (Zoopagales).*Mycologia* , *111* (2), 291–298. https://doi.org/10.1080/00275514.2018.1546066

De Filippis, F., Laiola, M., Blaiotta, G., & Ercolini, D. (2017). Different amplicon targets for sequencing-based studies of fungal diversity. *Applied and Environmental Microbiology* , *83* (17), 1–9. https://doi.org/10.1128/AEM.00905-17

de Mendiburu, F. (2020). *Statistical Procedures for Agricultural Research Package "agricolae"* (1.3-3). The R Project for Statistical Computing. https://cran.r-project.org/package=agricolae

Drechsler, C. (1938). New Zoopagaceae Capturing and Consuming Soil Amoebae. *Mycologia* , *30* (2), 137–157.

Duddington, C. L. (1955). Fungi That Attack Microscopic Animals.*Botanical Review* , *21* (7), 377–439.

Dutton, C. M., Christine, P., & Sommer, S. S. (1993). General method for amplifying regions of very high G + C content. Nucleic Acids Research, 21(12), 2953–2954. https://doi.org/10.1093/nar/21.12.2953

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* ,*32* (5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* , *26* (19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

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

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 Journal* , *4* (5), 642–647. https://doi.org/10.1038/ismej.2009.153

Feibelman, T., Bayman, P., & Cibula, W. G. (1994). Length variation in the internal transcribed spacer of ribosomal DNA in chanterelles.*Mycological Research* , *98* (6), 614–618. https://doi.org/10.1016/S0953-7562(09)80407-3

Furneaux, B., Bahram, M., Rosling, A., Yorou, N.S., Ryberg, M. (2021). Long- and short-read metabarcoding technologies reveal similar spatio-temporal structures in fungal communities. *Authorea* , doi: 10.22541/au.160340221.18389016/v1

Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology* , *2* (2), 113–118. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x

Gardes, M., & Bruns, T. D. (1996). Community strucutre of ectomycorrhizal fungi in a *Pinus muricata* forest: above- and below-ground views. *Canadian Journal of Botany* , *74* , 1572–1583.

Gazis, R., Rehner, S., & Chaverri, P. (2011). Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Molecular Ecology* , *20* (14), 3001–3013. https://doi.org/10.1111/j.1365-294X.2011.05110.x

Glass, D. J., Takebayashi, N., Olson, L. E., & Taylor, D. L. (2013). Evaluation of the authenticity of a highly novel environmental sequence from boreal forest soil using ribosomal RNA secondary structure modeling. *Molecular Phylogenetics and Evolution* , *67* (1), 234–245. https://doi.org/10.1016/j.ympev.2013.01.018

Gottlieb, A. M., & Lichtwardt, R. W. (2001). Molecular variation within and among species of Harpellales. *Mycologia* , *93* (1), 66–81. https://doi.org/10.1080/00275514.2001.12061280

Grossart, H. P., Wurzbacher, C., James, T. Y., & Kagami, M. (2016). Discovery of dark matter fungi in aquatic ecosystems demands a reappraisal of the phylogeny and ecology of zoosporic fungi.*Fungal Ecology* , *19* , 28–38. https://doi.org/10.1016/j.funeco.2015.06.004

Hart, M. M., Aleklett, K., Chagnon, P. L., Egan, C., Ghignone, S., Helgason, T., Lekberg, Y., Opik, M., Pickles, B. J., & Waller, L. (2015). Navigating the labyrinth: A guide to sequence-based, community ecology of arbuscular mycorrhizal fungi. *New Phytologist* ,*207* (1), 235–247. https://doi.org/10.1111/nph.13340

Hofstetter, V., Buyck, B., Eyssartier, G., Schnee, S., & Gindro, K. (2019). The unbearable lightness of sequenced-based identification.*Fungal Diversity* , *96* , 243–284. https://doi.org/https://doi.org/10.1007/s13225-019-00428-3 The

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* , *6* (2), 65–70.

Hopple, J. S., & Vilgalys, R. (1994). Phylogenetic Relationships among Coprinoid Taxa and Allies Based on Data from Restriction Site Mapping of Nuclear rDNA. *Mycologia* , *86* (1), 96–107.

Hsieh TC, Ma KH, Chao A (2020). iNEXT: Interpolation and Extrapolation for Species Diversity. R package version 2.0.20, http://chao.stat.nthu.edu.tw/wordpress/software_download/.

Huson, D., Mitra, S., & Ruscheweyh, H. (2011). MEGAN4 paper.*Genome Research* , *21* (9), 1552–1560. https://doi.org/10.1101/gr.120618.111.Freely

Ihrmark, K., Bodeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandstrom-Durling, M., Clemmensen, K. E., & Lindahl, B. D. (2012). New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* , *82* (3), 666–677. https://doi.org/10.1111/j.1574-6941.2012.01437.x

Jackson, H.S., & Dearden, E.R. (1948). *Martensella corticii*Thaxter and its distribution. *Mycologia, 40* (2), 168-176.

Johansen, R. B., Johnston, P., Mieczkowski, P., Perry, G. L. W., Robeson, M. S., Burns, B. R., & Vilgalys, R. (2016). A native and an invasive dune grass share similar, patchily distributed, root-associated fungal communities. *Fungal Ecology* , *23* , 141–155. https://doi.org/10.1016/j.funeco.2016.08.003

Jusino, M. A., Banik, M. T., Palmer, J. M., Wray, A. K., Xiao, L., Pelton, E., Barber, J. R., Kawahara, A. Y., Gratton, C., Peery, M. Z., & Lindner, D. L. (2019). An improved method for utilizing high-throughput amplicon sequencing to determine the diets of insectivorous animals. *Molecular Ecology Resources* , *19* (1), 176–190. https://doi.org/10.1111/1755-0998.12951

Koleff, P., Gaston, K. J., & Lennon, J. J. (2003). Measuring Beta Diversity for Presence-Absence Data. *Journal of Animal Ecology* ,*72* (3), 367–382.

Kurihara, Y., Degawa, Y., Tokumasu, S. (2001). A new genus*Myconymphaea* (Kickxellales) with peculiar septal plugs.*Mycological Research, 105* (11), 1397-1402.

Kurihara, Y., Sukarno, N., Ilyas, M., Yuniarti, E., Mangunwardoyo, W., Park, J-Y., Saraswati, R., Widyastuti, Y., Ando, K. (2008). Indonesian Kickxellales: two species of *Coemansia* and *Linderina* .*Mycoscience, 49* , 250-257.

Lazarus, K. L., Benny, G. L., Ho, H. M., & Smith, M. E. (2017). Phylogenetic systematics of *Syncephalis* (Zoopagales, Zoopagomycotina), a genus of ubiquitous mycoparasites. *Mycologia* ,*109* (2), 333–349. https://doi.org/10.1080/00275514.2017.1307005

Lazarus, K. L., & James, T. Y. (2015). Surveying the biodiversity of the Cryptomycota using a targeted PCR approach. *Fungal Ecology* ,*14* , 62–70. https://doi.org/10.1016/j.funeco.2014.11.004

Letcher, P.M., Powell, M.J. (2019). A taxonomic summary of Aphelidiaceae. *IMA Fungus* , 10(4), 1-11. https://doi.org/10.1186/s43008-019-0005-7

Li, S., Deng, Y., Wang, Z., Zhang, Z., Kong, X., Zhou, W., Yi, Y., & Qu, Y. (2020). Exploring the accuracy of amplicon-based internal transcribed spacer markers for a fungal community. *Molecular Ecology Resources* , *20* (1), 170–184. https://doi.org/10.1111/1755-0998.13097

Li, Y., Steenwyk, J.L., Chang, Y., Wang, Y., James, T.Y., Stajich, J.E., Spatafora, J.W., Groenewald, M., Dunn, C.W., Hittinger, C.T., Shen, X-X., Rokas, A. (2021). A genome-scale phylogeny of Fungi; insights into early evolution, radiations, and the relationship between taxonomy and phylogeny, bioRxiv preprint https://doi.org/10.1101/2020.08.23.262857

Lichtwardt, R. W. 1986. *The Trichomycetes: Fungal Associates of Arthropods* . Springer-Verlag, New York. 343 pp.

Linder, D. H. (1943). The genera *Kickxella, Martensella, and Coemansia* . *Farlowia* , *1* (1), 49–77. https://doi.org/10.2307/2420811

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* (4), 731–740. https://doi.org/10.3852/10-331

Liu, K. L., Porras-Alfaro, A., Kuske, C. R., Eichorst, S. A., & Xie, G. (2012). Accurate, rapid taxonomic classification of fungal large-subunit rRNA Genes. *Applied and Environmental Microbiology* , *78* (5), 1523–1533. https://doi.org/10.1128/AEM.06826-11

Lloyd-MacGilp, S. A., Chambers, S. M., Dodd, J. C., Fitter, A. H., Walker, C., & Young, J. P. W. (1996). Diversity of the ribosomal internal transcribed spacers within and among isolates of *Glomus mosseae* and related mycorrhizal fungi. *New Phytologist* ,*133* (1), 103–111. https://doi.org/10.1111/j.1469-8137.1996.tb04346.x

Lofgren, L. A., Uehling, J. K., Branco, S., Bruns, T. D., Martin, F., & Kennedy, P. G. (2019). Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles. *Molecular Ecology* , *28* (4), 721–730. https://doi.org/10.1111/mec.14995

Lucking, R., Aime, M. C., Robbertse, B., Miller, A., Ariyawansa, H., Aoki, T., Cardinali, G., Crous, P., Druzhinina, I., Geiser, D., Hawksworth, D., Hyde, K., Irinyi, L., Jeewon, R., Johnston, P., Kirk, P., Malosso, E., May, T., Meyer, W., ... Schoch, C. (2020). Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus* , *11* (1). https://doi.org/10.1186/s43008-020-00033-z

Maddison, W. P., & Maddison, D. R. (2019). *Mesquite Version 3.61* . http://www.mesquiteproject.org

Malar, C.M., Kruger, M., Kruger, C., Wang, Y., Stajich, J.E., Keller, J., Chen, E.C.H., Yildirir, G., Villeneuve-Laroche, M., Roux, C., Delaux, P.-M., Corradi, N. (2021). The genome of *Geosiphon pyriformis* reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology* , 31, 1-8. https://doi.org/10.1016/j.cub.2021.01.058

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods* , *7* (2), 111–118. https://doi.org/10.1038/nmeth.1419

Manter, D. K., & Vivanco, J. M. (2007). Use of the ITS primers, ITS1F and ITS4, to characterize fungal abundance and diversity in mixed-template samples by qPCR and length heterogeneity analysis. *Journal of Microbiological Methods* , *71* (1), 7–14. https://doi.org/10.1016/j.mimet.2007.06.016

McMurdie, P.J., & Holmes, S. (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE, 8* (4), e61217

Mota-Gutierrez, J., Ferrocino, I., Rantsiou, K., & Cocolin, L. (2019). Metataxonomic comparison between internal transcribed spacer and 26S ribosomal large subunit (LSU) rDNA gene. *International Journal of Food Microbiology* , *290* (October 2018), 132–140. https://doi.org/10.1016/j.ijfoodmicro.2018.10.010

Mueller, R. C., Belnap, J., & Kuske, C. R. (2015). Soil bacterial and fungal community responses to nitrogen addition across soil depth and microhabitat in an arid shrubland. *Frontiers in Microbiology* ,*6* (SEP). https://doi.org/10.3389/fmicb.2015.00891

Mueller, R. C., Gallegos-Graves, L. V., & Kuske, C. R. (2016). A new fungal large subunit ribosomal RNA primer for high-throughput sequencing surveys. *FEMS Microbiology Ecology* , *92* (2), 1–11. https://doi.org/10.1093/femsec/fiv153

Munch, K., Boomsma, W., Huelsenbeck, J. P., Willerslev, E., & Nielsen, R. (2008). Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology* , *57* (5), 750–757. https://doi.org/10.1080/10635150802422316

Nelson, J., & Shaw, A. J. (2019). Exploring the natural microbiome of the model liverwort: fungal endophyte diversity in *Marchantia polymorpha* L. *Symbiosis* , *78* (1), 45–59. https://doi.org/10.1007/s13199-019-00597-4

Nguyen, N. H., Smith, D., Peay, K., & Kennedy, P. (2015). Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist* , *205* (4), 1389–1393. https://doi.org/10.1111/nph.12923

Nilsson, R. Henrik, Ryberg, M., Kristiansson, E., Abarenkov, K., Larsson, K. H., & Koljalg, U. (2006). Taxonomic reliability of DNA sequences in public sequences databases: A fungal perspective. *PLoS ONE* , *1* (1). https://doi.org/10.1371/journal.pone.0000059

Nilsson, R. Henrik, Tedersoo, L., Abarenkov, K., Ryberg, M., Kristiansson, E., Hartmann, M., Schoch, C. L., Nylander, J. A. A., Bergsten, J., Porter, T. M., Jumpponen, A., Vaishampayan, P., Ovaskainen, O., Hallenberg, N., Bengtsson-Palme, J., Eriksson, K. M., Larsson, K.-H., Larsson, E., & Koljalg, U. (2012). Five

simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys* , *4* , 37–63. https://doi.org/10.3897/mycokeys.4.3606

Nilsson, R. Henrik, Wurzbacher, C., Bahram, M., Coimbra, V. R. M., Larsson, E., Tedersoo, L., Eriksson, J., Ritter, C. D., Svantesson, S., Sanchez-Garcia, M., Ryberg, M., Kristiansson, E., & Abarenkov, K. (2016). Top 50 most wanted fungi. *MycoKeys* , *12* , 29–40. https://doi.org/10.3897/mycokeys.12.7553

Nilsson, Rolf Henrik, 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 Research* ,*47* (D1), D259–D264. https://doi.org/10.1093/nar/gky1022

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *Community Ecology Package ' vegan '* (2.5-6). The R Project for Statistical Computing. https://cran.r-project.org

Opik, M., Davison, J., Moora, M., & Zobel, M. (2014). DNA-based detection and identification of Glomeromycota: The virtual taxonomy of environmental sequences. *Botany* , *92* (2), 135–147. https://doi.org/10.1139/cjb-2013-0110

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* , *2018* (5). https://doi.org/10.7717/peerj.4925

Perez-Izquierdo, L., Morin, E., Maurice, J. P., Martin, F., Rincon, A., & Buee, M. (2017). A new promising phylogenetic marker to study the diversity of fungal communities: The Glycoside Hydrolase 63 gene.*Molecular Ecology Resources* , *17* (6), e1–e11. https://doi.org/10.1111/1755-0998.12678

Picard, K. T. (2017). Coastal marine habitats harbor novel early-diverging fungal diversity. *Fungal Ecology* , *25* , 1–13. https://doi.org/10.1016/j.funeco.2016.10.006

Porras-Alfaro, A., Liu, K. L., Kuske, C. R., & Xiec, G. (2014). From genus to phylum: Large-subunit and internal transcribed spacer rRNA operon regions show similar classification accuracies influenced by database composition. *Applied and Environmental Microbiology* ,*80* (3), 829–840. https://doi.org/10.1128/AEM.02894-13

Porter, T. M., & Golding, G. B. (2012). Factors that affect large subunit ribosomal DNA amplicon sequencing studies of fungal communities: Classification method, primer choice, and error. *PLoS ONE* ,*7* (4). https://doi.org/10.1371/journal.pone.0035749

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools.*Nucleic Acids Research* , *41* (D1), 590–596. https://doi.org/10.1093/nar/gks1219

R Core Team. (2019). *R: A Language and Environment for Statistical Computing* (3.6.2). https://www.r-project.org/

Rambaut, A. (2012). *FigTree v1.4.3* . Institute of Evolutionary Biology, University of Edinburgh. http://tree.bio.ed.ac.uk/software/figtree

Reynolds, N. K., Benny, G. L., Ho, H.-M., Hou, Y.-H., Crous, P. W., & Smith, M. E. (2019). Phylogenetic and morphological analyses of the mycoparasitic genus *Piptocephalis* . *Mycologia* ,*111* (1). https://doi.org/10.1080/00275514.2018.1538439

Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C., & Jaffe, D. B. (2013). Characterizing and measuring bias in sequence data. *Genome Biology* , *14* (5). https://doi.org/10.1186/gb-2013-14-5-r51

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., ... Schindel, D. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America* ,*109* (16), 6241–6246. https://doi.org/10.1073/pnas.1117018109

Skelton, J., Jusino, M. A., Carlson, P. S., Smith, K., Banik, M. T., Lindner, D. L., Palmer, J. M., & Hulcr, J. (2019). Relationships among wood-boring beetles, fungi, and the decomposition of forest biomass.*Molecular Ecology* , *28* (22), 4971–4986. https://doi.org/10.1111/mec.15263

Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., James, T. Y., O'Donnell, K., Roberson, R. W., Taylor, T. N., Uehling, J., Vilgalys, R., White, M. M., & Stajich, J. E. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* , *108* (5), 1028–1046. https://doi.org/10.3852/16-042

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* ,*30* (9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

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., ... Robert, V. (2015). One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes.*Persoonia: Molecular Phylogeny and Evolution of Fungi* ,*35* (1), 242–263. https://doi.org/10.3767/003158515X689135

Taylor, D. L., Walters, W. A., Lennon, N. J., Bochicchio, J., Krohn, A., Caporaso, J. G., & Pennanen, T. (2016). Accurate Estimation of Fungal Diversity and Abundance through Improved Lineage-Specific Primers Optimized for Illumina Amplicon Sequencing. *Applied and Environmental Microbiology* , *82* (24), 7217–7226. https://doi.org/http://dx.doi.org/10.1128 /AEM.02576-16

Tedersoo, L., Anslan, S., Bahram, M., Koljalg, U., & Abarenkov, K. (2020). Identifying the 'unidentified' fungi: a global-scale long-read third-generation sequencing approach. *Fungal Diversity* ,*103* (1), 273–293. https://doi.org/10.1007/s13225-020-00456-4

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

Tedersoo, L., Bahram, M., Puusepp, R., Nilsson, R. H., & James, T. Y. (2017). Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome* , *5* (1), 1–10. https://doi.org/10.1186/s40168-017-0259-5

Tedersoo, L., & Lindahl, B. (2016). Fungal identification biases in microbiome projects. *Environmental Microbiology Reports* ,*8* (5), 774–779. https://doi.org/10.1111/1758-2229.12438

Tedersoo, L., Tooming-Klunderud, A., Anslan, S. (2017b). PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and perspectives. *New Phytologist, 217* , 1370-1385. doi: 10.1111/nph.14776

Thiery, O., Moora, M., Vasar, M., Zobel, M., & Opik, M. (2012). Inter- and intrasporal nuclear ribosomal gene sequence variation within one isolate of arbuscular mycorrhizal fungus, *Diversispora* sp.*Symbiosis* , *58* (1–3), 135–147. https://doi.org/10.1007/s13199-012-0212-0

Toju, H., Tanabe, A. S., Yamamoto, S., & Sato, H. (2012). High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE* , *7* (7). https://doi.org/10.1371/journal.pone.0040863

Torres-Cruz, T. J., Tobias, T. L. B., Almatruk, M., Hesse, C. N., Kuske, C. R., Desiro, A., Benucci, G. M. N., Bonito, G., Stajich, J. E., Dunlap, C., Arnold, A. E., & Porras-Alfaro, A. (2017).*Bifiguratus adelaidae*

, gen. et sp. nov., a new member of Mucoromycotina in endophytic and soil-dwelling habitats.*Mycologia* , *109* (3), 363–378. https://doi.org/10.1080/00275514.2017.1364958

Truong, C., Gabbarini, L. A., Corrales, A., Mujic, A. B., Escobar, J. M., Moretto, A., & Smith, M. E. (2019). Ectomycorrhizal fungi and soil enzymes exhibit contrasting patterns along elevation gradients in southern Patagonia. *New Phytologist* , *222* (4), 1936–1950. https://doi.org/10.1111/nph.15714

Vandepol, N., Liber, J., Desiro, A., Na, H., Kennedy, M., Barry, K., Grigoriev, I. V., Miller, A. N., O'Donnell, K., Stajich, J. E., & Bonito, G. (2020). Resolving the Mortierellaceae phylogeny through synthesis of multigene phylogenetics and phylogenomics. *Fungal Diversity* , *In Press* .

Větrovský, T., Kolařík, M., Žifčáková, L., Zelenka, T., & Baldrian, P. (2016). The rpb2 gene represents a viable alternative molecular marker for the analysis of environmental fungal communities. *Molecular Ecology Resources* , *16* (2), 388–401. https://doi.org/10.1111/1755-0998.12456

Vilgalys, R., & Hester, M. (1990). Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several*Cryptococcus* species. *Journal of Bacteriology* ,*172* (8), 4238–4246. https://doi.org/10.1128/jb.172.8.4238-4246.1990

Vu, D., Groenewald, M., Vries, M. De, Gehrmann, T., Stielow, B., Eberhardt, U., Groenewald, J. Z., Cardinali, G., Houbraken, J., Boekhout, T., Crous, P. W., & Robert, V. (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. *Studies in Mycology* , *154* , 135–154. https://doi.org/10.1016/j.simyco.2018.05.001

Walsh, E., Luo, J., Khiste, S., Scalera, A., Sajjad, S., & Zhang, N. (2020). Pygmaeomycetaceae, a new root associated family in Mucoromycotina from the pygmy pine plains. *BioRxiv* , 2020.07.03.187096. https://doi.org/10.1101/2020.07.03.187096

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* ,*73* (16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wang, Y., White, M.M., Moncalvo, J-M. (2019). Diversification of the gut fungi Smittium and allies (Harpellales) co-occurred with the origin of complete metamorphosis of their symbiotic insect hosts (lower Diptera). Molecular Phylogenetics and Evolution, 139, 1-13. https://doi.org/10.1016/j.ympev.2019.106550

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

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* . Springer-Verlag. https://ggplot2.tidyverse.org

Xue, C., Hao, Y., Pu, X., Ryan Penton, C., Wang, Q., Zhao, M., Zhang, B., Ran, W., Huang, Q., Shen, Q., & Tiedje, J. M. (2019). Effect of LSU and ITS genetic markers and reference databases on analyses of fungal communities. *Biology and Fertility of Soils* , *55* (1), 79–88. https://doi.org/10.1007/s00374-018-1331-4

Yahr, R., Schoch, C. L., & Dentinger, B. T. M. (2016). Scaling up discovery of hidden diversity in fungi: Impacts of barcoding approaches.*Philosophical Transactions of the Royal Society B: Biological Sciences* , *371* (1702). https://doi.org/10.1098/rstb.2015.0336

[dataset] Reynolds, N.K., Jusino, M.A., Stajich, J.E., Smith, M.E.; 2020; OSF (https://osf.io/cz3mh/); doi:

Figure 1. Violin plots of the average percentage of forward and reverse reads that were merged to form contigs from each primer set (ITS1F/ITS2, LROR/LR3 and LR22F/LR3). The number of total reads returned for each dataset is listed above the boxes and the number of total (i.e. fungal and non-fungal) OTUs found after filtering and quality control is given below.

Figure 2. Primer set variation (ITS1F/ITS2, LROR/LR3, LR22F/LR3) in OTU length and read number assigned to each OTU according to taxonomic assignment by Kingdom and fungal phylum using the hybrid (for ITS1F) or UTAX (for LSU) method (A, B) and the RDP classifier (C, D). Box plot height and whiskers represent OTU length range, whereas the box plot width represents the proportion of reads assigned to each group. The LROR primer set was almost entirely comprised of forward reads with a length of 283 bp, resulting in a flat line. SUPP FIG 4 contains additional graphs of the LROR data separately. Percentages indicate the proportion of non-fungal (A, C) and fungal (B, D) OTUs in each dataset. Note that Glomeromycotina is a subphylum within the Mucoromycota but is categorized separately for comparison with the ITS taxonomy.

Figure 3. Relative abundance of reads assigned to each phylum of early diverging fungi by sample type for each primer set.

Figure 4. Non-metric multi-dimensional scaling (NMDS) ordination plots for all fungi and early diverging fungal communities recovered by the primer sets (ITS1F/ITS2, LROR/LR3, LR22F/LR3) for all California (CA) and Florida (FL) environmental sampling sites. Point colors represent different sampling locations and point shapes indicate primer set. Stress values are listed for each dataset.

Figure 5. Comparison of alpha diversity measures for early diverging fungi for each primer set by site (colors) and sample type (shapes).

Figure 6. Maximum likelihood phylogenetic reconstruction of fungal LSU sequences including references from GenBank and newly sequenced isolates from this study. 50 OTUs identified only as "Fungi" from each of the LROR and LR22F datasets were included and the numbers are bolded. Analyses were performed in RAxML v 8 using the GTR + GAMMA model and 1,000 bootstrap replicates. Classes of fungi are colored if they include unknown LSU OTUs or shaded grey if they do not. The dark grey "BLAST match to protists" shading indicates that these clades are comprised of OTUs that had GenBank matches to protist sequences with the OTU IDs in red. Asterisks indicate early diverging clades. SUPP TABLE 5 has a list of all OTUs included in the phylogeny along with their BLAST matches.

Figure 7. Maximum likelihood phylogenetic reconstruction of Zoopagomycota LSU sequences including references from GenBank and newly sequenced isolates from this study. OTUs identified as Zoopagomycota species from each of the LROR and LR22F datasets were included and the numbers are bolded and include the order to which each OTU was classified. Analyses were performed in RAxML v 8 using the GTR + GAMMA model and 1,000 bootstrap replicates. Branch supports [?]70 are shown.

Table 1. Zoopagomycota mock community members and results of mock community OTU taxonomy comparisons across the ITS1F/ITS2, LROR/LR3, and LR22F/LR3 primer sets, including the target fragment length in base pairs (bp), GenBank accession numbers (ITS, LSU), and the primer columns list the number of OTUs of each mock community member detected by the RDP classifier taxonomy (outside parentheses) versus the RDP+SILVA LSU database (for LSU) or the AMPtk hybrid method (for ITS1F) (in parentheses).

| Isolate $^{+}$ | ITS1F/ITS2 length bp $^{++\S}$ | LR22F/LR3 length |
|---|---|---|
| *Coemansia aciculifera* **NRRL 2694** | 412 | 365 |
| *Coemansia thaxteri* **IMI 214463** | 412 | 363$^{\hat{}}$ |
| *Piptocephalis cruciata* **CBS 826.97** | 527 | 446 |
| *Piptocephalis microcephala* **CBS 418.77** | 457 | 437 |
| *Syncephalis californica* **A23985** | 433$^{+}$ | 503$^{+}$ |
| *Syncephalis pseudoplumigaleata* **S71** | 778$^{\hat{}}$ | 503 |
| *Coemansia erecta* **IMI 312319** | 287$^{+}$ | 361$^{+}$ |
| *Coemansia interrupta* **BCRC 34489** | 216$^{\hat{}}$ | 241 |
| *Dimargaris xerosporica* **NRRL 3178** | 728$^{+}$ | 261$^{+}$ |
| *Piptocephalis cylindrospora* **RSA 2659** | 427 | 455 |
| *Piptocephalis moniliformis* **NRRL 13723** | 335$^{\hat{}}$ | 468 |
| *Syncephalis cornu* **NRRL A-5447 (61)** | 407$^{+}$ | 501 |

| Isolate [+] | ITS1F/ITS2 length bp [++§] | LR22F/LR3 length |
|---|---|---|
| *Syncephalis depressa* **S116 (4)** | 271ˆ | 491 |
| *Tieghemiomyces parasiticus* **RSA 861** | Unknown | 227 |
| *Acaulopage tetraceros* **T2** | 281 | 416 |
| *Coemansia* **sp. RSA 1933** | 219 | |
| *Coemansia* **sp. RSA 2604** | 186 | |
| *Piptocephalis graefenhanii* **S99022101** | 280[+] | 522[+] |
| *Syncephalis digitata* **S521 (K12)** | 270 | 458 |
| *Syncephalis obconica* **S227 (K10)** | 321 | 458 |
| *Mycoemilia scoparia* **NBRC 100468** | Unknown | 354 |
| *Myconymphaea yatsukahoi* **NBRC 100467** | Unknown | 370 |
| *Pinnaticoemansia coronantispora* **CBS 131509** | Unknown | 370ˆ |
| *Stylopage hadra* **B35** | 483[+] | 421 |
| *Zoophagus pectospora* **B39** | Unknown | 354 |
| *Acaulopage acanthospora* **Ac1** | Unknown | 416 |
| *Zoopage* **sp. C3** | 326 | 461 |
| *Acaulopage* **sp. Ap** | 246ˆ | 416 |
| *Cochlonema odontospora* **E1** | Unknown | 389 |
| *Zoopage* **sp. Zo2** | Unknown | 461 |
| **Hosts (*Cokeromyces, Cunninghamella, Rhizopus, Umbelopsis*)** | Typically short (˜300) | |
| Unidentified Kickxellales | Typically short (˜300) | |
| Non-zygomycetes (animals, protists, Dikarya fungi) | | |
| "Unclassified" or "Fungi" | | |

[+]Isolate names in bold indicate species for which genomic DNAs were combined to make the mock community. The remaining "isolates" refer to OTU identifications that were recovered after amplification, sequencing, and bioinformatic processing.[++]Lengths with ˆ indicate that the number is an estimate due to the reference sequence missing one or more priming sites. Lengths with [+] indicate that the length is an estimate based on a different isolate of the same species, with the exception of *Dimargaris xerosporica* for which estimates are based on the only species with available rDNA sequences, *D. bacillispora* . [¶]Bolded GenBank accession numbers indicate isolates newly sequenced for this study.

Table 2. Results of PERMANOVA and whether ANOVA analyses of the*B* sim output were significant for all fungi and early diverging fungal (EDF) communities recovered by each primer set (ITS1F/ITS2, LROR/LR3, LR22F/LR3) by sites within states (California, CA and Florida, FL) and by sample type (including invertebrates, mud, soil, and water), as well as the interaction between site and sample type, with asterisks indicating significant p-values (<0.05).

| PERMANOVA | Degrees of freedom | Sum squares | $R^2$ | F value | P value | ANOVA *B*sim significant? |
|---|---|---|---|---|---|---|
| ITS1F CA fungi Site | 4 | 9.845 | 0.21154 | 8.8969 | 0.0001* | no |
| ITS1F CA fungi Sample | 3 | 3.755 | 0.08068 | 4.5242 | 0.0001* | yes |
| ITS1F CA fungi Site x Sample | 11 | 9.149 | 0.19658 | 3.0065 | 0.0001* | |
| ITS1F FL fungi Site | 1 | 1.3797 | 0.0855 | 4.5509 | 0.0001* | yes |

| PERMANOVA | Degrees of freedom | Sum squares | $R^2$ | F value | P value | ANOVA *B*sim significant? |
|---|---|---|---|---|---|---|
| ITS1F FL fungi Sample | 3 | 3.4 | 0.2107 | 3.7382 | 0.0001* | yes |
| ITS1F FL fungi Site x Sample | 3 | 2.2622 | 0.14018 | 2.4872 | 0.0001* | |
| ITS1F CA EDF Site | 4 | 9.678 | 0.19172 | 7.4519 | 0.0001* | yes |
| ITS1F CA EDF Sample | 3 | 3.784 | 0.07495 | 3.8844 | 0.0001* | no |
| ITS1F CA EDF Site x Sample | 11 | 9.746 | 0.19306 | 2.7287 | 0.0001* | |
| ITS1F FL EDF Site | 1 | 1.4805 | 0.093 | 4.7734 | 0.0001* | no |
| ITS1F FL EDF Sample | 3 | 3.8367 | 0.241 | 4.1235 | 0.0001* | yes |
| ITS1F FL EDF Site x Sample | 3 | 1.9184 | 0.1205 | 2.0617 | 0.0007* | |
| LROR CA fungi Site | 4 | 9.328 | 0.18756 | 7.6923 | 0.0001* | yes |
| LROR CA fungi Sample | 3 | 5.096 | 0.10247 | 5.6033 | 0.0001* | yes |
| LROR CA fungi Site x Sample | 11 | 9.542 | 0.19185 | 2.8612 | 0.0001* | |
| LROR FL fungi Site | 1 | 1.2176 | 0.08762 | 4.4197 | 0.0001* | no |
| LROR FL fungi Sample | 3 | 3.6358 | 0.26163 | 4.399 | 0.0001* | no |
| LROR FL fungi Site x Sample | 3 | 2.1557 | 0.15512 | 2.6082 | 0.0001* | |
| LROR CA EDF Site | 4 | 8.618 | 0.1528 | 5.4378 | 0.0001* | yes |
| LROR CA EDF Sample | 3 | 5.205 | 0.09228 | 4.3787 | 0.0001* | yes |
| LROR CA EDF Site x Sample | 11 | 8.899 | 0.15779 | 2.0418 | 0.0001* | |
| LROR FL EDF Site | 1 | 1.2958 | 0.12837 | 6.4159 | 0.0001* | no |
| LROR FL EDF Sample | 3 | 3.4713 | 0.34388 | 5.7291 | 0.0001* | yes |

| PERMANOVA | Degrees of freedom | Sum squares | $R^2$ | F value | P value | ANOVA *B*sim significant? |
|---|---|---|---|---|---|---|
| LROR FL EDF Site x Sample | 3 | 1.692 | 0.16761 | 2.7925 | 0.0001* | |
| LR22F CA fungi Site | 4 | 9.531 | 0.2103 | 9.1561 | 0.0001* | yes |
| LR22F CA fungi Sample | 3 | 4.397 | 0.09702 | 5.6324 | 0.0001* | yes |
| LR22F CA fungi Site x Sample | 11 | 9.274 | 0.20461 | 3.2395 | 0.0001* | |
| LR22F FL fungi Site | 1 | 1.32 | 0.0919 | 5.1877 | 0.0001* | no |
| LR22F FL fungi Sample | 3 | 3.7993 | 0.26452 | 4.9772 | 0.0001* | yes |
| LR22F FL fungi Site x Sample | 3 | 2.1189 | 0.14753 | 2.7758 | 0.0001* | |
| LR22F CA EDF Site | 4 | 9.335 | 0.16055 | 5.7059 | 0.0001* | yes |
| LR22F CA EDF Sample | 3 | 4.364 | 0.07506 | 3.5568 | 0.0001* | yes |
| LR22F CA EDF Site x Sample | 11 | 9.68 | 0.16648 | 2.1516 | 0.0001* | |
| LR22F FL EDF Site | 1 | 1.3695 | 0.10074 | 6.0836 | 0.0001* | no |
| LR22F FL EDF Sample | 3 | 4.1174 | 0.30287 | 6.0965 | 0.0001* | no |
| LR22F FL EDF Site x Sample | 3 | 2.2544 | 0.16583 | 3.3381 | 0.0001* | |

Table 3. Mantel test results for the correlation analyses between the distance matrices of each primer set (ITS1F/ITS2, LROR/LR3, LR22F/LR3) for fungal and early diverging fungal (EDF) OTUs using Spearman's rank correlation, asterisks indicate significant p-values (<0.05).

| Comparison | R statistic | Significance$^+$ | 90% quantile | 95% quantile | 97.5% quantile | 99% quantile |
|---|---|---|---|---|---|---|
| ITS1F vs LROR EDF | 0.5397 | 0.006* | 0.0345 | 0.0436 | 0.0539 | 0.0641 |
| ITS1F vs LR22 EDF | 0.5603 | 0.006* | 0.0318 | 0.0408 | 0.0487 | 0.0543 |
| LROR vs LR22F EDF | 0.5849 | 0.006* | 0.0356 | 0.0483 | 0.0598 | 0.0698 |
| ITS1F vs LROR fungi | 0.4716 | 0.006* | 0.0389 | 0.0532 | 0.0637 | 0.0740 |
| ITS1F vs LR22F fungi | 0.7101 | 0.006* | 0.0373 | 0.0481 | 0.0545 | 0.0667 |
| LROR vs LR22F fungi | 0.4483 | 0.006* | 0.0424 | 0.0518 | 0.0601 | 0.0771 |

[+]P-values adjusted for multiple comparisons using the Holm method (Holm, 1979).