12-4-2023

Evaluating the Effectiveness of the Internal Transcribed Sequences (ITS) as DNA Barcodes to Estimate Fungal Diversity

Jailisa Linares
Nova Southeastern University

Capstone of
Jailisa Linares

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
Biological Sciences

Nova Southeastern University
Halmos College of Arts and Sciences

December 2023

Approved:
Capstone Committee

Committee Chair: Dr. Navdeep Gill

Committee Member: Dr. Matthew Johnston

Committee Member: Dr. Christopher Blanar

NOVA SOUTHEASTERN UNIVERSITY
HALMOS COLLEGE OF ARTS AND SCIENCES

Evaluating the Effectiveness of the Internal Transcribed Sequences (ITS) as DNA Barcodes to Estimate Fungal Diversity

By

Jailisa Linares

Submitted to the Faculty of
Halmos College of Arts and Sciences
in partial fulfillment of the requirements for
the degree of Master of Science with a specialty in:

Biological Sciences

Nova Southeastern University
January 2024

Abstract

Universal phylogenetic markers such as the nuclear ribosomal internal transcribed sequences (ITS), specifically ITS1 and ITS2, are routinely used to estimate fungal diversity in environmental samples. However, numerous studies report differences in the performance and efficacy of ITS1 and ITS2 in documenting fungal diversity. To better understand the implications of using ITS1 versus ITS2, a comprehensive representation of the diverse fungal taxa was necessary to conduct a meta-analysis of their use across multiple fungal taxa. In order to address this, a thorough literature review was conducted to compare and contrast the use of ITS1 and ITS2 as effective DNA barcodes. Publicly available datasets were used to synthesize a simulated fungal community representing diverse fungal taxonomic groups and the efficacy of the two amplicons was tested and compared to the complete ITS. This study hypothesized that ITS1 and ITS2 are not equally effective for the resolution of fungal taxa. Specifically, when comparing ITS1 and ITS2 for phylogenetic resolution, an overlapping set of taxa were identified by both approaches, whereas certain taxa were preferentially resolved by individual ITS amplicons. The evaluation presented here should allow readers to develop a better understanding of the uses and limitations of ITS1 versus ITS2 in studying fungal diversity and ecology and enable them to develop improved approaches for better taxonomic resolution as well as aid in identification of potentially novel species.

Key words: fungal diversity, metagenomics, internal transcribed sequences: ITS1, ITS2, fungal barcoding, phylogenetics, evolution, fungal pathogens, long read sequencing, primers, fungal genomics, variations.

Acknowledgements

This study and all that has come along with it has been a very rewarding and valuable experience. I would like to thank my major advisor Dr. Navdeep Gill, for her continuing support and dedication to making this a success. I'd like to extend a thank you to Dr. Emily Schmitt for her openness and willingness to appropriately tend to all needs associated with this study and my success in the program. A big thank you to Dr. Christopher Blanar for his open communication and invaluable support to me and the entire Biology Department. Thank you to Dr. Matthew Johnston for his support and feedback that made this a successful review. I'd also like to thank Dr. Braham Dhillon at the University of Florida for his invaluable efforts with the data collection and analysis associated with this study. All of your help has been greatly appreciated. Lastly, thank you to my amazing family, friends and mentors who continue to guide and ground me as I find my way through academia.

Table of Contents

I. Introduction *1*

I.1. Fungal Taxonomy 1

I.1.1 Why is it important to study fungal diversity 2

I.1.2 Molecular markers to study fungal taxonomy 3

I.2. What is ITS? 4

I.3. ITS as markers in fungi 4

I.4. DNA barcoding using ITS for metagenomics 6

I.5. ITS1 vs ITS2 6

II. Statement of Purpose *7*

III. Materials & Methods *8*

III.1. Literature Review 8

III.2. Data Collection 9

III.2.1. Compiling a database of complete fungal ITS sequences 9

III.2.2 Collection of primer sequences for ITS1, ITS2 and complete ITS region 10

III.3. Data analysis 12

III.4. Data visualization 14

IV. Results *17*

IV.1. Data structure and pre-processing 17

IV.2. AMPtk pipeline 18

IV.3. Taxonomic assignments for ITS, ITS1 and ITS2 from AMPtk output 19

IV.4. Taxonomic Resolution at the Phyla and Class Level using Phyloseq 21

V. Discussion *29*

V.1. Future prospects 33

VI. Conclusions *34*

VII. References *35*

I. Introduction

I.1. Fungal Taxonomy

Fungi are heterotrophic eukaryotic organisms found in most habitats on Earth, with nearly 150,000 described species. Fungi play a critical role in nutrient cycling and decomposition, act as pathogens and parasites, and form symbiotic relationships with plants, animals, and algae (Naranjo-Ortiz and Gabaldon 2019). Fungal taxonomy is the scientific study of how fungi are classified, named, and organized into a hierarchical system. Biologists have used a range of characteristics to classify fungi, including their physical and biochemical features, as well as their genetic relationships. At the highest level, fungi are divided into several phyla based on their morphology and genetic relationships (Cavalier-Smith 1998; Guarro et al. 1999). The classification of fungi has constantly evolved as new information, and in that evolution, has become available through advances in molecular biology and other scientific fields (Tkacz et al. 2004). By exploring the diversity and the evolution of fungi, we can attribute practical applications such as identifying fungal pathogens and developing strategies for controlling fungal diseases in agriculture and medicine. The most significant major phyla which represented the greatest diversity in this study included Ascomycota, Basidiomycota, Zygomycota, Chytridiomycota, and Glomeromycota (Table 1).

Table 1. Classification of Fungi. Phyla and classifications of fungi based on their morphological differences with examples of each assigned.

Phylum	**Characteristics**	**Examples**
Ascomycota	Produce spores in sac-like structures called asci	Yeasts, molds, truffles, morels
Basidiomycota	Produce spores on club-shaped structures called basidia	Mushrooms, bracket fungi, rusts
Zygomycota	Produce spores in protective structures called zygosporangia	Bread molds, *Pilobolus*
Chytridiomycota	Produce flagellated spores	Aquatic fungi, *Batrachochytrium dendrobatidis*
Glomeromycota	Form arbuscular mycorrhizae with plant roots	*Endogone*, *Glomus*

I.1.1 Why is it important to study fungal diversity

Fungal taxonomy and phylogeny play an important role in the study of fungal diversity because it provides a systematic framework for understanding the evolutionary relationships between different fungal species and groups. Traditionally, fungal taxonomy was based on several morphological characters, such as colony morphology, pigmentation, spore bearing structures both asexual and sexual, and spore shape, size and attachment. Classification of fungi into taxonomic categories allows biologists to identify patterns of diversity and relatedness among diverse fungal species and generate a deeper understanding of the ecological roles and potential applications of different fungi (Wu et al. 2019). Fungal taxonomy has proven important for practical applications, such as agriculture, medicine, and biotechnology (Tkacz et al. 2004). Next generation sequencing technologies have accelerated the pace of discovery in fungal taxonomy, as evidenced from genome sequencing initiatives in academia, that provide an opportunity to significantly increase the number of sequenced fungal genomes in the public domain (Tkacz et al. 2004). The study of fungal biodiversity and their genomes is imperative for the evolution of fungal ecology (Despres-Loustau et al. 2007). Generation of sequence

data contributes to the development and improvement of bioinformatic tools needed to reliably annotate fungal genomes and conduct comparative genomics inquiries related to genome and species evolution. Deciphering the genomes of fungi also allows for the development of agricultural fungicides and therapeutic agents for human and animal health (Tkacz et al. 2004). The interaction of fungal species with humans, plants, and animals ranges across the spectrum of beneficial to harmful and the knowledge gained from studying fungal biology and diversity helps in developing strategies for managing fungal diseases, identifying new drug targets, and harnessing the biotechnological potential of fungi.

I.1.2 Molecular markers to study fungal taxonomy

Morphological traits are not always reliable when establishing fungal phylogenies: some characters such as spore shape, size and pigmentation exhibit phenotypic plasticity, varying in relation to nutrient availability; fungi that do not produce asexual or sexual spores may not have discriminating morphological characters for taxonomic placement. Processes such as hybridization and convergent evolution may lead to development of similar morphological characters in evolutionary unrelated fungi, thereby complicating their taxonomic placement. In such cases, standard DNA based molecular markers are needed for taxonomic resolution. These molecular markers are referred to as “barcodes”; they are DNA sequence codes for “core” proteins that are present in most or all organisms, displaying mutation rates that exceed speciation rates in most lineages, and are easy to amplify (Stoeckle and Hebert 2008). Several genes, such as ITS, SSU, LSU, RPB1, RPB2, EF1a, MCM7, and COI have been identified that fit these criteria, but the most commonly used DNA barcode in fungal phylogenetics is the ITS region (Raja et al. 2017; Gutierrez et al. 2010).

I.2. What is ITS?

Internal Transcribed Sequences (ITS) are regions of eukaryotic DNA located between the 18S and 28S ribosomal RNA genes containing conserved and variable regions that can be amplified using polymerase chain reaction (PCR) and sequenced to determine the identity of the organism (Figure 1) (Reller et al. 2007). They are commonly used in molecular biology as markers for species identification and classification, making them extremely useful for phylogenetic analysis. Amplification of the conserved and variable ITS regions of the fungal genome allows for easy identification of fungal species.

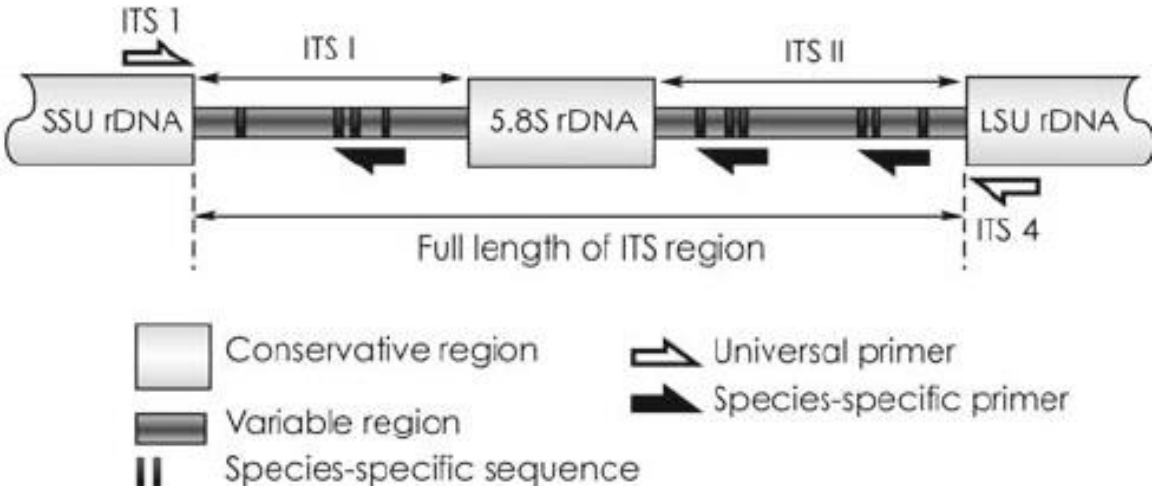

Figure 1. ITS regions – conserved and variable regions identified. Adapted from "Qualitative and quantitative PCR methods using species-specific primer for detection and identification of wood rot fungi," by S.S Horisawa and Y.S. Doi, 2009, *Journal of Wood Science, 55*, p. 135. Copyright 2009 by the Japan Wood Research Society. Adapted with permission.

I.3. ITS as markers in fungi

Given the extensive diversity of fungi and the novelty of DNA barcoding, the adoption of ITS markers in fungi faced several hurdles. Among those, it was important to consider variability and available molecular data. The variability of ITS sequences among

different species is due to the accumulation of mutations over time. This variability has provided an opportunity to develop a more novel species recognition system. In this study, this challenge is undertaken by using ITS markers to identify fungal species and differentiate them into specific phyla. There are five preliminary steps to produce a fungal barcode using ITS (Table 2). Additionally, considering molecular data available has become an invaluable step in the identification and studying of taxa and phylogeny due to the rapidly growing diversity of fungi and the taxing effort of keeping up with this data via publicly available metagenomic datasets (Tekpinar and Kalmer 2019).

DNA barcoding using ITS refers to the use of internal transcribed sequence regions in ribosomal DNA as a tool for identifying and classifying fungal species (Vu et al. 2009). The ITS region is commonly used as a barcode because it contains both conserved and variable region sequences that can be used to differentiate between fungal species based on their unique ITS sequence; this unique ITS sequence is the most commonly used marker for identification of fungal species due to its high variability and resolution (Christ et al. 2011).

Table 2. Fundamental steps of fungal barcoding using ITS. Identification of fungal species requires the retrieval and analysis of a sample to be determined. These steps are described in detail below, alongside the context in which it was needed for this study.

Fungal barcoding steps	Significance for fungal barcoding using ITS
1. Sample Collection	Fungal samples are collected from the environment, such as soil or plant tissues.
2. DNA Extraction	DNA is extracted from fungal samples using standard molecular biology techniques.
3. PCR Amplification	ITS region from ribosomal DNA are amplified using polymerase chain reaction (PCR) with specific primers
4. Sequence Analysis	The amplified ITS regions are sequenced and compared to reference sequences in databases, such as nt database (NCBI), to identify species.
5. Phylogenetic Analysis	The sequences are used for phylogenetic analysis to infer the evolutionary relationships among different fungal species.

I.4. DNA barcoding using ITS for metagenomics

The utility of DNA barcodes goes beyond identifying individual fungi collected and isolating them from environmental samples. These barcodes can be used to analyze entire fungal communities derived from environmental samples and help understand ecological questions related to fungal community structure and species richness in varied ecosystems. DNA extracted from environmental samples is amplified using barcoding primers, generating sequence data for thousands of fungi, collectively known as metagenome. The ITS region in fungi ranges from 600-900 bp (Toju et al. 2012), so the complete region cannot be amplified routinely using NGS platforms such as Illumina due to the limitation of long read sequencing. Therefore, researchers need to choose between ITS1 and ITS2 when conducting metagenomic studies.

I.5. ITS1 vs ITS2

In metagenomic datasets, studies have reported the preferential use of ITS1, ITS2 or alternately the complete ITS sequence. A large-scale meta-analysis of ITS1 and ITS2

regions showed that the ITS1 region significantly provided more resolution than ITS2 regions for ascomycete fungi (Wang et al. 2014). The ITS1 region may be a better choice than ITS2 due to its higher level of sequence conservation, making it more reliable for species identification and delimitation (Hollingsworth et al. 2011). Species delimitation in ITS1 and ITS2 is used to distinguish between closely related species and determine their taxonomic boundaries (Nolasco et al. 2022). An increase in the use of molecular techniques to investigate the evolutionary history of species would help resolve issues related to species delimitation (Harrup et al. 2015). In contrast, ITS2 has been reported to be a better choice in some applications because it is more variable and can provide higher resolution for phylogenetic analysis (Wolf et al. 2013). In terms of species richness, Bazzicalupo et al. (2013) favors the use of ITS2 due to larger variability and a higher recovery of operational taxonomic units.

The use of ITS1 and ITS2 as DNA barcodes have their own strengths and weaknesses and the best choice will depend on the specific research question and the fungal community being investigated. It's also possible that a combination of both ITS1 and ITS2 may provide the best results, but the large read size of the next generation sequencing methods in available metagenomic datasets are a limitation.

II. Statement of Purpose

The objective of this study was to evaluate the effectiveness of three amplicon sequences derived from full-length ITS regions (ITS1, ITS2, and complete ITS) as DNA barcodes to estimate fungal diversity. Initially, a survey of existing research on the use of ITS amplicons for fungal metagenomics was done in order to understand the current status of ITS1 and ITS2 in capturing fungal diversity. Numerous studies have examined

fungal community structure across varied environmental samples using different sequencing technologies and data analysis methods. This study hypothesized that ITS1 and ITS2 were not equally effective for the resolution of fungal taxa. Full-length ITS sequences obtained from reference sequence database were split into ITS1 and ITS2 regions and characterized using the AMPtk data analysis pipeline. The hypothesis was then tested by evaluating the effectiveness of ITS1, ITS2 and the complete ITS region, in capturing the maximum fungal diversity in a simulated fungal community of 15,000+ fungal species.

III. Materials & Methods

III.1. Literature Review

DNA barcoding has made leaps in the scientific community from its early emergence in the 1900's to the development of DNA based markers and PCR technological advancements available today (Yasui 2020). These technologies, in addition to accessibility to open resources like the Fungal Genome Project (FGP), have increased support for claims stating that DNA markers like ITS1 and ITS2 could serve as the future of DNA barcoding (Bazzicalupo et al. 2013). With many different ongoing projects in support of this argument, articles for this review were selected using databases like ScienceDirect, PubMed, ProQuest etc. This study was supported by articles published within the last 30 years, ranging from January 1993 to January 2023. This timeframe allowed for a comprehensive study from the early emergence of DNA based technologies and research, to what is currently acceptable in terms of DNA based approaches for studying and identifying fungal diversity.

The selection of articles throughout this investigation was appropriate to research previously published in areas associated with fungal genomics and fungal barcoding using markers such as ITS1 and ITS2. Some keywords used in these searches included but were not limited to: fungal barcoding, DNA markers: ITS1 and ITS2, fungal taxonomy, phylogenetics, fungal genomics and invasive pathogens. Research articles containing metagenomic datasets and traditional sanger sequencing datasets were selected for review, and comparisons were made on how the use of ITS as DNA markers has changed with the evolution of Next Generation Sequencing techniques (NGS).

III.2. Data Collection

III.2.1. Compiling a database of complete fungal ITS sequences

A total of 15,921 full-length ITS sequences were obtained from the National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq) Targeted Loci Project online database. Nucleotide sequences were obtained from two bioprojects that are related to fungi and oomycetes:

1. Bioproject 177353: Fungal Internal Transcribed Spacer RNA (ITS) RefSeq Targeted Loci Project.
2. Bioproject 362621: Oomycetes Internal Transcribed Spacer RNA (ITS) RefSeq Targeted Loci Project].

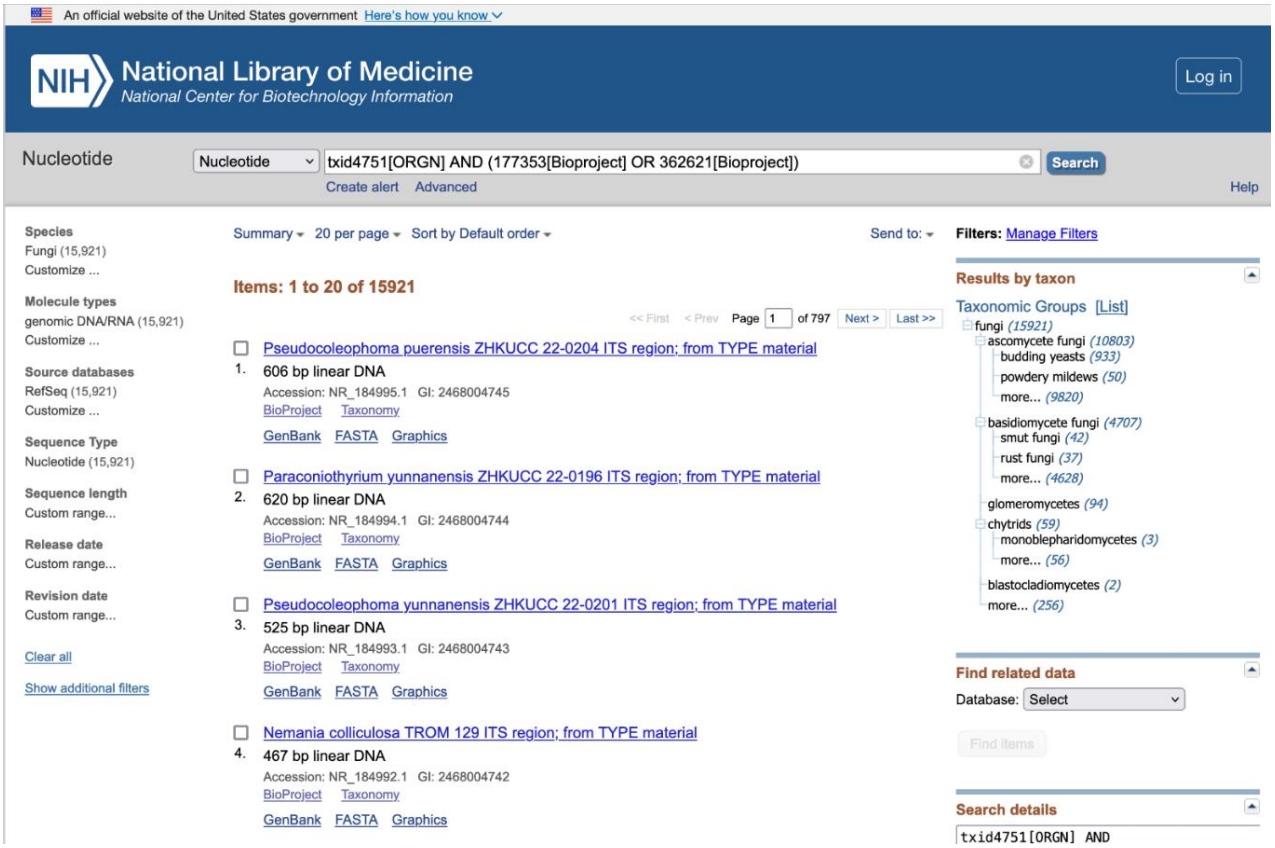

Figure 2. Screenshot showing the 15,921 fungal ITS sequences download from NCBI. Complete sequences were retrieved and downloaded from the NCBI database. All 15,921 nucleotide sequences are publicly available. Two bioprojects were used: Bioproject 177353, Fungal Internal Transcribed Spacer RNA (ITS) RefSeq Targeted Loci Project and, Bioproject 362621, Oomycetes Internal Transcribed Spacer RNA (ITS) RefSeq Targeted Loci Project.

III.2.2 Collection of primer sequences for ITS1, ITS2 and complete ITS region

Primer sequences for ITS1, ITS2 and complete ITS region were obtained from four sources (Gardes et al. 1993; White et al. 1990; Martin et al. 2005; Blaalid et al. 2013). Primers listed in Table 3 have been shown to successfully amplify the complete ITS region from entire fungal genomes and have also been used to effectively amplify ITS1 and ITS2.

Table 3. Primer sequences for the complete ITS region, ITS1, and ITS2. Primer sequences listed were used to determine the ITS region of interest needed for identification of fungi. Primers mentioned served as a blueprint for ITS amplification. For more complete representation an additional primer sequence was also given but may or may not have been used for analysis (Gardes et al. 1993; White et al. 1990; Martin et al. 2005; Blaalid et al. 2013).

ITS	Primer sequences	Supplemental primer sequences
Complete ITS region	**ITS1F:** CTTGGTCATTTAGAGGAAGTAA (White et al. 1990)	**ITS1:** CCGTAGGTGAACCTGCGG (Gardes and Bruns, 1993)
	ITS4: TCCTCCGCTTATTGATATGC (White et al. 1990)	**ITS4:** TCCTCCGCTTATTGATATGC (White et al. 1990)
ITS1	**ITS1F:** CTTGGTCATTTAGAGGAAGTAA (White et al. 1990)	**ITS1F:** TCCGTAGGTGAACCTGCGG (Gardes and Bruns, 1993)
	ITS2R: GCTGCGTTCTTCATCGATGC (White et al. 1990)	**ITS5:** GGAAGTAAAAGTCGTAACAAGG (White et al. 1990)
ITS2	**ITS3:** GCATCGATGAAGAACGCAGC (White et al. 1990)	**ITS3:** GCATCGATGAAGAACGCAGC (White et al. 1990)
	ITS4: TCCTCCGCTTATTGATATGC (White et al. 1990)	**ITS4-B:** TCCTCCGCTTATTGATATGC (White et al. 1990)

The placement of different primer sets relative to each other along with their respective amplicons are shown in Figure 3. The primer information displayed was used in the downstream data analysis pipeline.

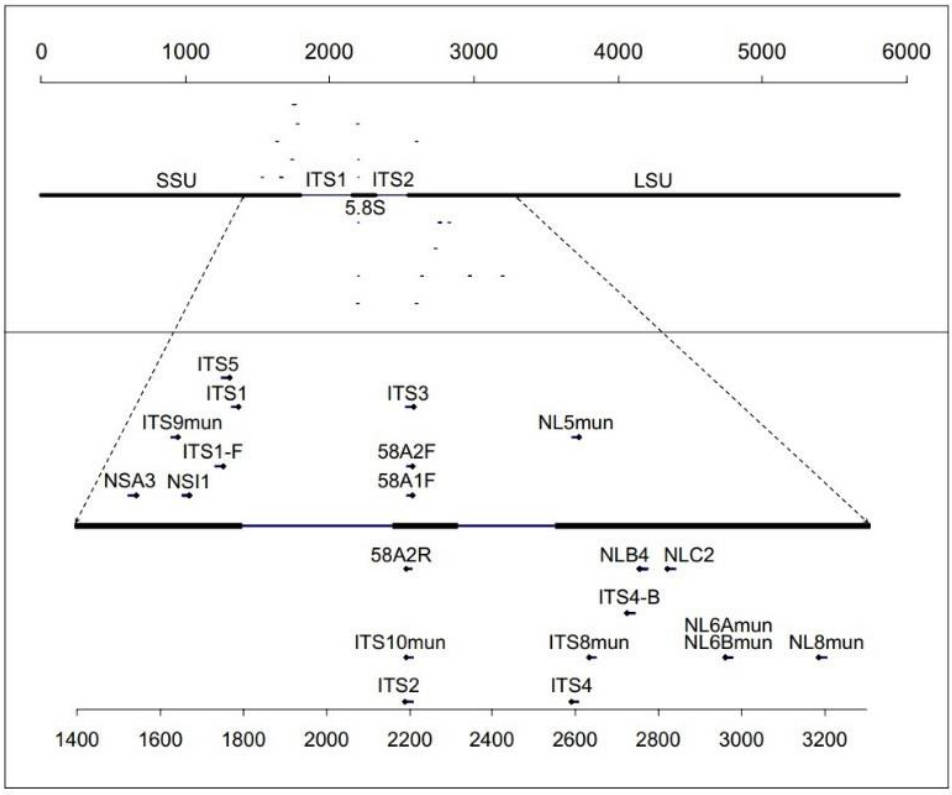

Figure 3. Primer sequences for extraction of specific ITS regions. Forward and reverse primers for ITS, ITS1 and ITS2 sequences are depicted. Adapted from "Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts," by K. J. Martin and P.T. Rygiewicz, 2005, *BMC Microbiology, 5*, p.3. Copyright 2005 by BioMed Central Ltd. Adapted with permission.

III.3. Data analysis

ITSx (https://microbiology.se/software/itsx/) was used to split the complete ITS sequences into ITS1 and ITS2. ITSx is an open-source software for the detection and extraction of the ITS1 and ITS2 sequences from a given ribosomal ITS sequence (Bengtsson-Palme et al. 2013). The software is a commonly used tool to study fungal ITS and is especially used in environmental sequencing samples.

The three sets of amplicons were run through the AMPtk pipeline (Palmer et al. 2018) using a series of scripts specifically designed to process NGS amplicon data using

USEARCH (Edgar 2010) and VSEARCH (Rognes et al 2016). The AMPtk pipeline performs a global alignment to the eukaryotic ITS sequences in the reference UNITE database at a given threshold of 97% identity to return unambiguous OTU assignments. The AMPtk taxonomic assignment involves the four major steps as outlined below:

1. **Pre-processing reads**: The pre-processing code was necessary in the AMPtk program to analyze the raw reads that were obtained by the NCBI download. The finding and removal of artifacts such as, barcodes/indexes, and primers, and trimming the reads to desired length, was done in this step to ensure data quality.
2. **Clustering data**: The pre-processed data was clustered to obtain Operational Taxonomic Units (OTUs). AMPtk offers several clustering algorithms, like UPARSE, DADA2, UNOISE2, and UNOISE3, to cluster sequences based on pre-defined percent identity cut-off, ensuring that reads that align are included in the same cluster. then clustered those reads into the OTUs found in Table 4.
3. **Filtering data:** This step is useful to remove sequencing artifacts that result from a phenomenon known as 'barcode crossover', i.e., where a small percentage of reads maybe assigned to the wrong sample. It is done on an OTU-basis to ensure that low-abundance OTUs are not preferentially parsed out. The OTU table from the clustering step is used as an input for this step.
4. **Assigning taxonomy:** Lastly, taxonomic assignment is done by comparing representative sequences from each OTU to a pre-built reference taxonomy database. Several classifiers such as BLAST, SINTAX, UTAX, Global Alignment (USEARCH/VSEARCH) and RDP Classifier, are available in AMPtk to estimate taxonomy. AMPtk, by default, uses a conservative 'hybrid' approach to assign

taxonomy by calculating a consensus LCA (least common ancestor) taxonomy based on results from three classifiers.

The AMPtk pipeline is equipped with databases appropriately formatted for the analysis of fungal ITS, fungal LSU, bacterial 16S, and insect COI amplicons. The Operational Taxonomic Units (OTUs) provided by the AMPtk clustering were analyzed to determine the fungal diversity represented by each of the three sets of amplicons.

For each amplicon, the following sets of commands were used in the AMPtk pipeline.

1. *$ amptk illumina -i FASTQ -o mydata -f GGAAGTAAAAGTCGTAACAAGG -r ITS2 --require_primer on*
2. *$ amptk dada2 -i mydata.demux.fq.gz --platform illumina -o out_dada2*
3. *$ amptk taxonomy -f out_dada2.ASVs.fa -i out_dada2.otu_table.txt -m mydata.mapping_file.txt -d ITS1*
4. *$ amptk taxonomy -f out_dada2.cluster.otus.fa -i out_dada2.cluster.otu_table.txt -m mydata.mapping_file.txt -d ITS1*

III.4. Data visualization

OTUs obtained from the AMPtk pipeline were further processed via downstream tools such as Phyloseq (McMurdie and Holmes 2013) for interactive analysis and graphical display of the data. R Studio (www.rstudio.com) was used to run Phyloseq using the AMPtk output. The files that were used in R studio were “otu_table” and “otu_taxonomy” containing information on the number of fungal sequences present in each OTU (otu_table) and number of fungal sequences present in each taxonomic level

such as Kingdom, Phylum, Class, Order, Family, Genus, and Species (otu_taxonomy). The original OTU file was modified into a format that was compatible to Phyloseq analysis. Reads that were unknown at a given resolution level were labeled as such to avoid any discrepancies in R. The AMPtk output files complete with metadata and taxonomy assigned as OTUs, were directly imported into R package Phyloseq for further analysis and data visualization. The top 1,000 OTUs in terms of abundance and/or number of sequences/OTU were selected for Phyloseq. This criterion was used for clarity and to avoid over stacking of the bar plots. This approach provided sufficient results to show the differences in taxonomic resolution of all three amplicons at the phylum and class level.

For each amplicon, the following sets of commands were used for successful running of the Phyloseq pipeline. The focus was on the fungal phyla, so "phylum" was used as the taxonomic level in the Phyloseq commands listed below. Only the most abundant 1,000 OTUs were plotted to avoid stacking and excessive visuals in the R plots.

1. *$ fungi <- subset_taxa(physeq, Domain=="Fungi")*

The OTUs with the highest number of sequences was obtained by inputting command:

2. *$ TopNOTUs <- names(sort(taxa_sums(fungi), TRUE)[1:1000])*

Additionally, subsets for the most abundant fungal phyla present were generated with Phyloseq. These phyla were Ascomycota and Basidiomycota. The commands were as follows:

3. *$ asco <- subset_taxa(physeq, Phylum=="Ascomycota") TopNOTUs <- names(sort(taxa_sums(asco), TRUE)[1:1000])*
4. *$ basidio <- subset_taxa(physeq, Phylum=="Basidiomycota") TopNOTUs <- names(sort(taxa_sums(basidio), TRUE)[1:1000])*

As a result, three R plots per amplicon were produced containing the resolution of each ITS at the phylum and class level.

Venny 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/) is a publicly available resource that was used to generate a visual representation of the taxonomic assignments for the complete ITS fungal sequences, ITS1 and ITS2 from the AMPtk output. The venn diagram was generated to show which genus was best represented in the AMPtk pipeline for each amplicon.

The results from the different steps of the AMPtk pipeline are shown in Figure 4. To summarize, AMPtk performed a global alignment of the ITS source sequences (15,853 for each amplicon) to the reference database UNITE and found the best match. This program is designed to give each individual read their own reference number in that database. The program is designed to output with high accuracy a matching fungal sequence to the best hit in the reference database. The criteria was set to a 97% cut-off for best accuracy of matches.

IV. Results

IV.1. Data structure and pre-processing

NCBI was used to download a fungal ITS type-material reference dataset containing 15,921 DNA sequences. The complete dataset was 11 megabase pairs (Mbp) in size and comprised eight phyla, 55 classes, 201 orders, 647 families, 3199 genus and 15,758 species (Table 6). The two bioprojects were downloaded both as a summary file and a GenBank file for complete metadata. The complete ITS sequences were run through ITSx to detect and split the ITS region into ITS1 and ITS2. ITSx (Rivers 2018 and Johan Bengtsson-Palme et al. 2013) used the provided primers sets and returned a FASTQ file trimmed and merged into a log file for analysis. Some sequences from this ITSx could not be split as they did not meet the size criteria for the sequence split. 15,853 sequences out of 15,921 were found to have ITS2 using ITSx. 15,914 sequences out of 15,921 were found to have ITS1 using ITSx. 15,853 ITS sequences were common among all three amplicons. Thus only 15,853 sequences were successfully split into both ITS1 and ITS2 with the use of ITS source data of 15,921 reference ITS sequences. The resulting 15,853 sequences were chosen as the final set for further analysis using the AMPtk pipeline on the three amplicons.

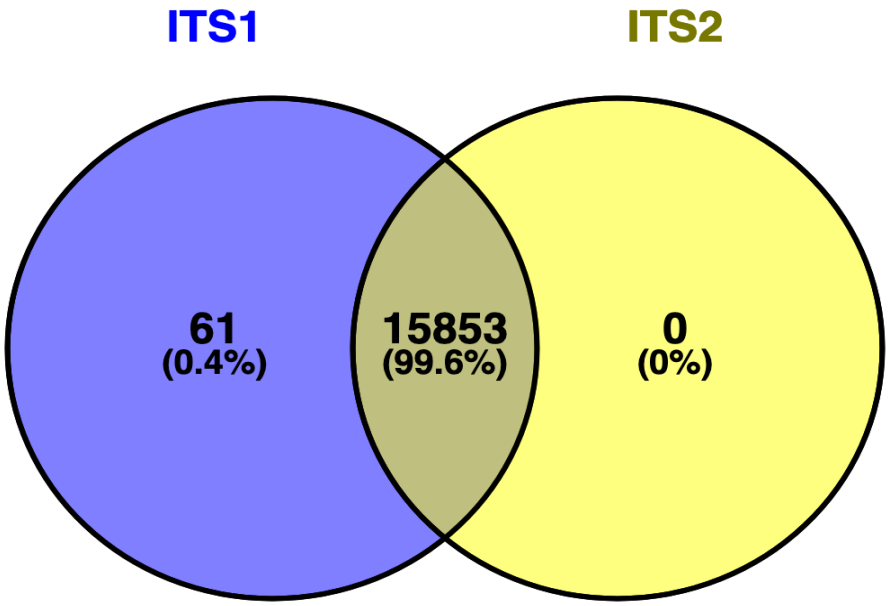

Figure 4. Venn diagram showing the results of ITSx split of complete ITS sequence into ITS1 and ITS2. A total of 15,853 were selected to represent all three amplicons.

IV.2. AMPtk pipeline

AMPtk was used to cluster the three amplicons into OTUs based on similarity to a UNITE database (https://unite.ut.ee/). Three AMPtk runs were performed i.e. one Amptk pipeline per amplicon was run with specific parameters and conditions as mentioned in the Materials and Methods section earlier.

OTUs were generated for the three amplicons. AMPtk was given 15,853 reads per sample run per amplicon (Figure 5). 15,768 sequences were filtered and retained for ITS; 15,265 for ITS1 and 15,335 for ITS2 based on the criteria that was mentioned in materials and methods. A final set of OTUs were generated as shown (Figure 5). The 15,853 sequences were assigned into a set of final OTUs for each amplicon. The ITS, ITS1 and ITS2 returned 7,948 (ITS), 7,930 (ITS1) and 7,693 (ITS2) OTUs, respectively. The number of sequences assigned to each OTU is shown in the labeled bars (Figure 5).

It should be noted that these numbers of final OTUs contain overlapping set of OTUs between the three amplicons, from which a non-redundant set of 9,664 OTUs was identified.

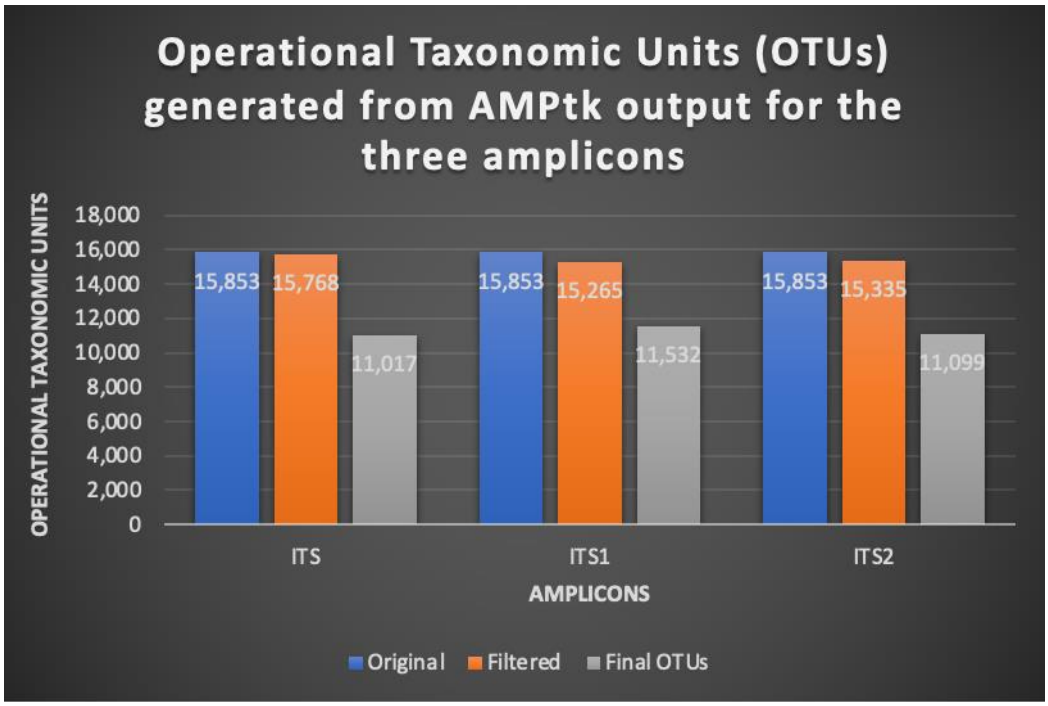

Figure 5. Operational Taxonomic Units (OTUs) generated from AMPtk output for ITS, ITS1 and ITS2. Each ITS displays three bar graphs. The first (blue) is representative of the total starting sequences of OTU's generated from the AMPtk output. The second (orange) represents the total filtered number of OTU's and the last (grey) demonstrates the final OTU AMPtk output.

IV.3. Taxonomic assignments for ITS, ITS1 and ITS2 from AMPtk output

The overview of taxonomic assignments obtained from the AMPtk output are shown in the corresponding Venn diagram (Fig. 6). Of the 15,853 sequences AMPtk identified a non-redundant set of 9,664 OTUs. Of these, 6,060 OTUs were shared amongst all three amplicons (Fig. 6). This implies ~63% of the OTUs were shared between all three amplicons, ITS, ITS1 and ITS2. 482 (5%) OTUs were only assigned a taxonomy by ITS, 698 (7.2%) OTUs were only resolved by ITS1, and 635 (6.3%) OTUs

were only resolved by ITS2 (Fig. 6). When comparing OTUs that were shared between two of the three amplicons, we observed that 791 OTUs overlapped between ITS and ITS1, 384 OTUs overlapped in ITS1 and ITS2, and 619 OTUs overlapped in ITS and ITS2 (Fig. 6).

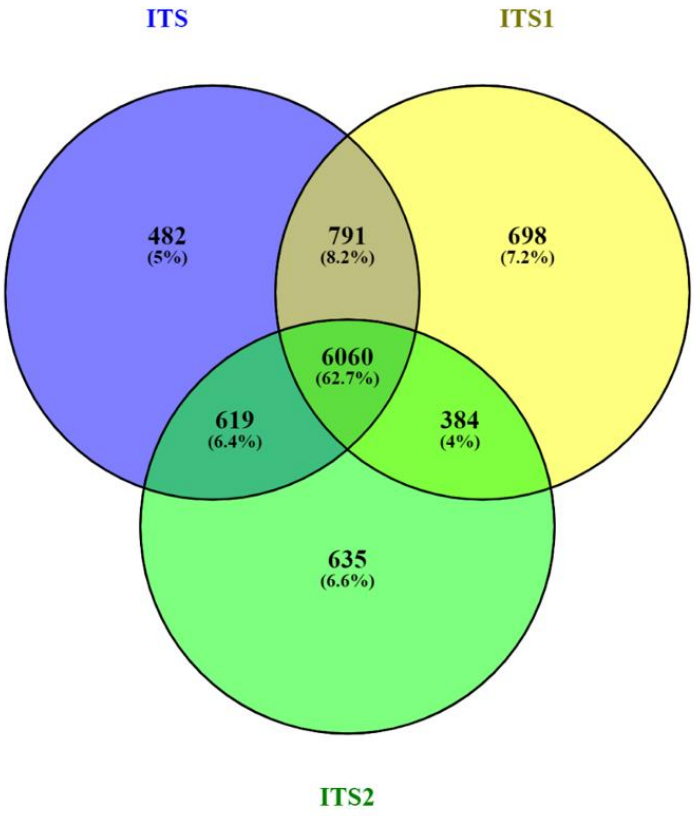

Figure 6. Taxonomic assignments for ITS, ITS1 and ITS2 resulting from AMPtk output. Individual taxonomy present in each amplicon varies and is illustrated above. This representation shows the taxonomic terms that are independent to and shared between amplicons.

The OTUs were assigned at seven different taxonomic levels represented by Kingdom/Domain (Level 1), Phylum (Level 2), Class (Level 3), Order (Level 4), Family (Level 5), Genus (Level 6), and Species (Level 7). Our results show different levels of taxonomic resolution. It is also important to note that some sequences were only resolved

to a genus level but not necessarily to species level for all OTUs, even though these sequences in the original dataset were derived from known species. There were some OTUs that were assigned to taxonomic groups outside of the Domain fungi, such as, unknowns, no hits, protozoa, and viridiplantae. These OTUs were excluded from further analysis.

IV.4. Taxonomic Resolution at the Phyla and Class Level using Phyloseq

The R plots show the taxonomic resolution using Phyloseq for the complete ITS region, ITS1 and ITS2 (Fig. 7, 8, 9). The taxonomic resolution is shown at two levels- Phyla (taxonomic level 2) and Class (taxonomic level 3) for each set of amplicon (Fig. 7, 8, 9). The results show a clear difference in the taxonomic classification of the fungal species at both these levels. At the phylum level, ITS and ITS1 captured more Basidiomycota than ITS2. Glomeromycota was captured the best by ITS relative to ITS1 and ITS2. Additionally, Glomeromycota was captured the least by ITS1. At the phylum level, ITS1 and ITS captured Ascomycota relatively similarly. At the phylum level, Basidiomycota was resolved the least by ITS2 (Fig. 9). The phylum Ascomycota was captured the best by ITS2 as compared to both ITS and ITS1.

The list of fungal classes that make up the most abundant OTUs vary between the three amplicons as is seen by the difference in the classes represented in the legends for ITS, ITS1 and ITS2 in both Ascomycota and Basidiomycota plots. It is worth noting that this pattern might not hold true when we are analyzing all the OTUs, not just the top 1,000.

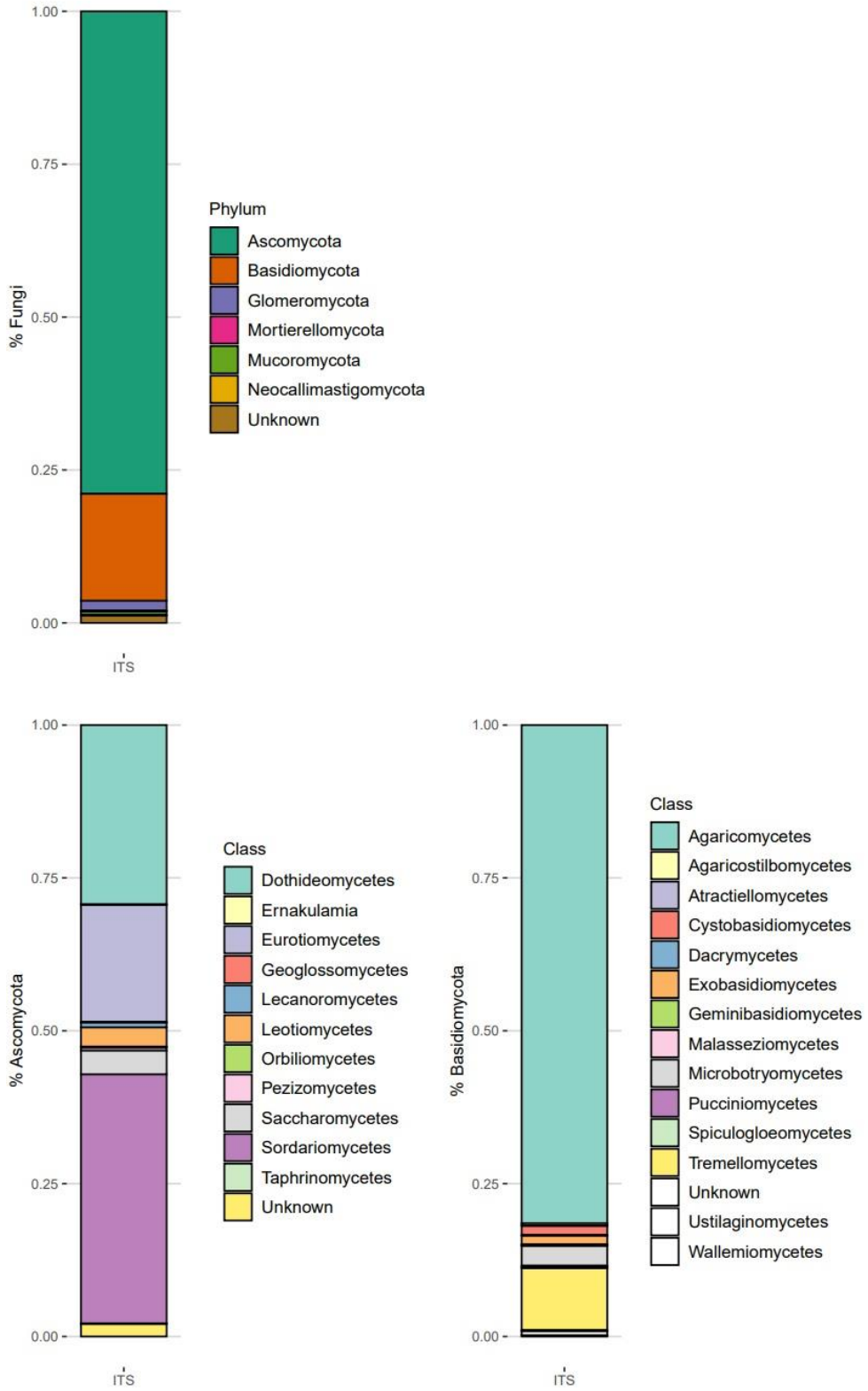

Figure 7. Taxonomic resolution by complete ITS using Phyloseq. Top 1000 OTUs are shown. For class level resolution, look at the legend not the color because colors change between each ITS amplicon.

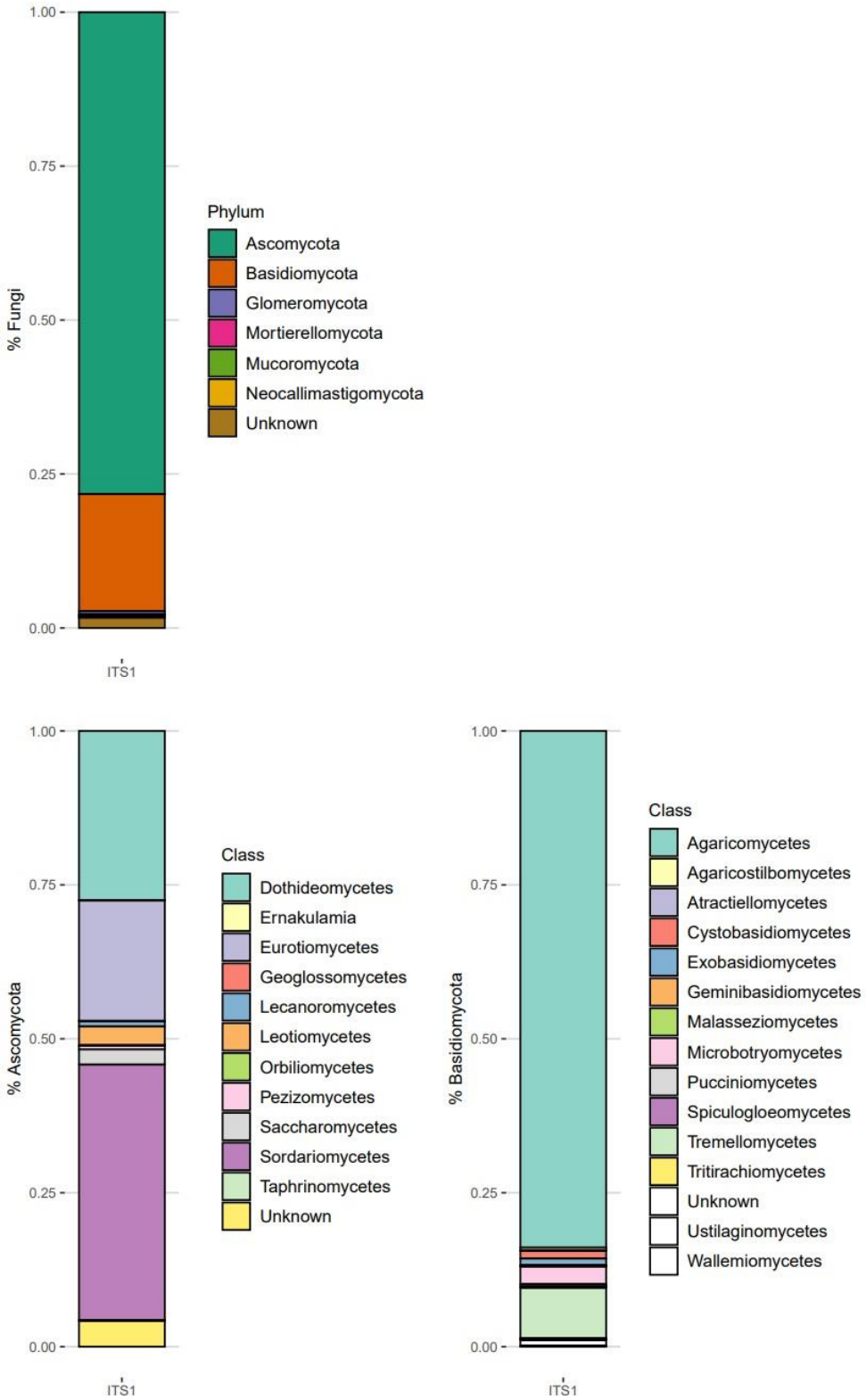

Figure 8. Taxonomic resolution by complete ITS using Phyloseq. Top 1000 OTUs are shown. For class level resolution, look at the legend not the color because colors change between each ITS amplicon.

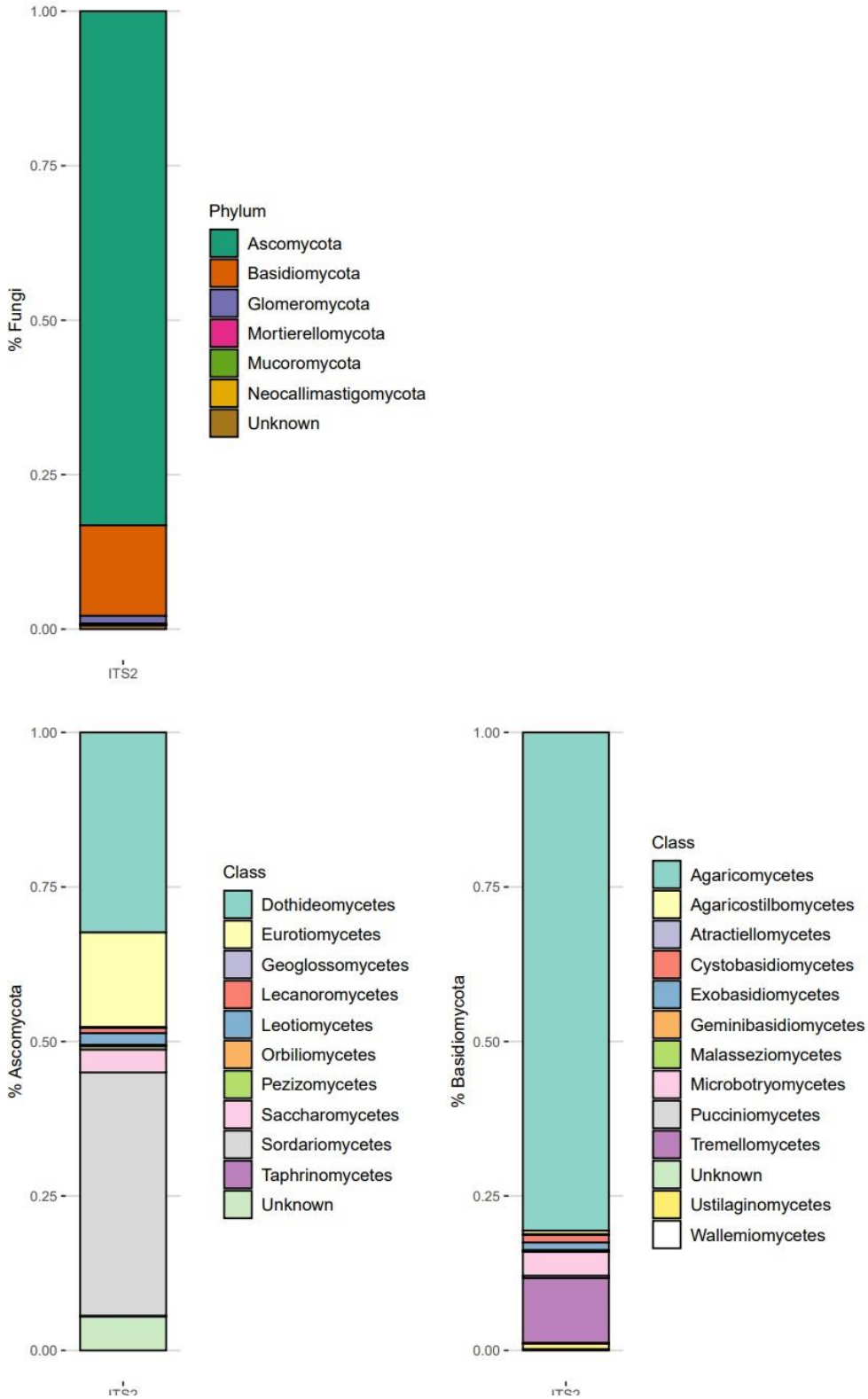

Figure 9. Taxonomic resolution by complete ITS using Phyloseq. Top 1000 OTUs are shown. For class level resolution, look at the legend not the color because colors change between each ITS amplicon.

To get an in-depth analysis of the taxonomic resolution at phylum and class levels, we analyzed all OTUs representing the 15 phyla and 56 classes comprising 14,871 (ITS), 14,118 (ITS1) and14,252 (ITS2) sequences, respectively. The number of fungal sequences that were assigned a taxonomic term at phylum level resolution is shown in Table 4.

Table 4. Taxonomic resolution in terms of number of sequences assigned to each phylum in the three amplicon sets. A total of 15 phyla comprising a total of 14,871 (ITS), 14,118 (ITS1) and 14,252 (ITS2) sequences, respectively, are shown.

	Phylum	**ITS**	**ITS1**	**ITS2**
1	Ascomycota	10,103	9,527	9,577
2	Basidiomycota	4,307	4,166	4,260
3	Mucoromycota	136	136	135
4	Unknown	111	133	74
5	Glomeromycota	92	33	88
6	Mortierellomycota	64	64	64
7	Chytridiomycota	29	30	26
8	Neocallimastigomycota	10	10	10
9	Kickxellomycota	7	7	7
10	Basidiobolomycota	4	4	3
11	Entorrhizomycota	2	2	2
12	Monoblepharomycota	2	2	2
13	Calcarisporiellomycota	1	1	1
14	Entomophthoromycota	1	1	1
15	Olpidiomycota	1	1	1
16	Rozellomycota	1	1	1

A comparison of the top four most abundant phyla namely, Ascomycota, Basidiomycota, Glomeromycota and Unknown (sequences without resolution) are shown in Figure 10.

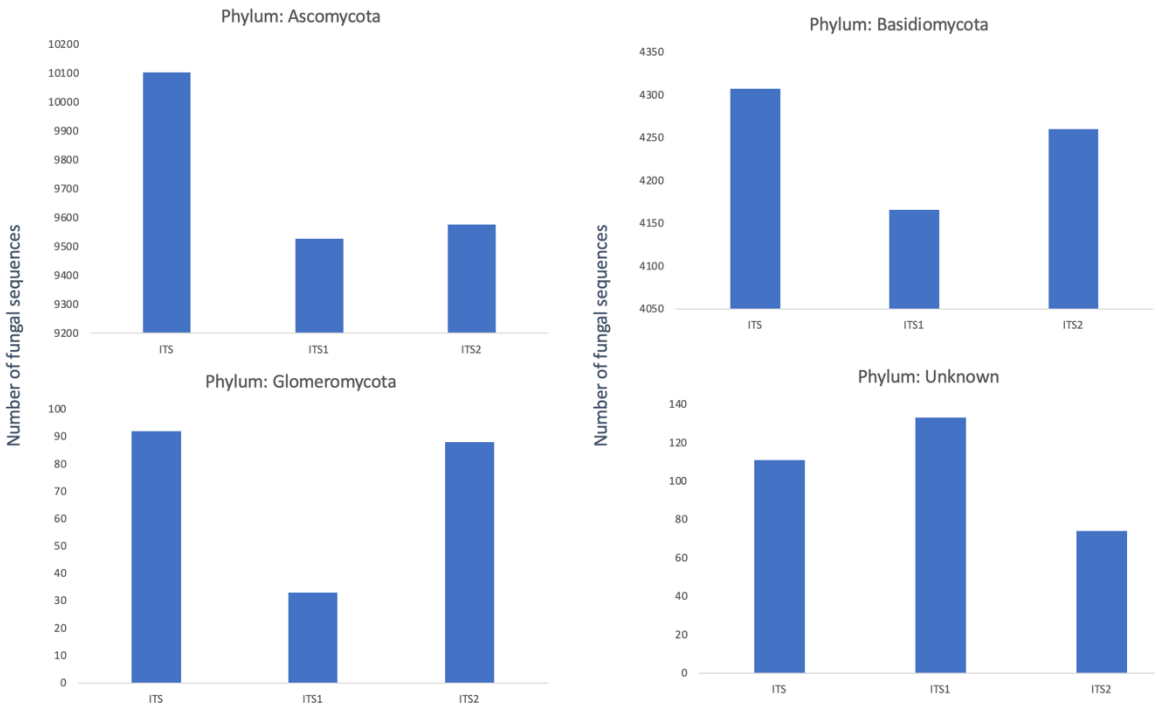

Figure 10. Comparison of the top most abundant phyla as resolved by each of the three amplicons, ITS, ITS1 and ITS2.

At the phylum level, ITS2 appears to be a better choice for taxonomic resolution as shown above (Table 4 and Fig. 10) especially for the phyla Basidiomycota and Glomeromycota. Moreover, fewer sequences remained unresolved in ITS2 as compared to ITS1 as seen above based on unknown phylum (Figure 10), also suggesting that ITS2 is better equipped for taxonomic resolution at the phylum level. Complete ITS, however performed better than both ITS1 and ITS2 in all the cases (Table 4).

Similarly, the number of fungal sequences that were assigned to 56 taxonomic classes is shown in Table 5. A total of 31 out of 56 taxonomic classes with fewer than 10 sequences per taxonomic class were merged for clarity.

Table 5. Taxonomic resolution in terms of number of sequences assigned to each class in the three amplicon sets. A total 56 taxonomic classes comprising a total of 14,871 (ITS), 14,118 (ITS1) and 14,252 (ITS2) sequences, respectively are shown. *31 taxonomic classes with less than 10 sequences per taxonomic class were grouped.

	Class	**ITS**	**ITS1**	**ITS2**
1	Sordariomycetes	3,586	3,454	3,388
2	Agaricomycetes	3,516	3,439	3,481
3	Dothideomycetes	2,813	2,650	2,758
4	Eurotiomycetes	1,536	1,435	1,327
5	Saccharomycetes	864	651	713
6	Leotiomycetes	465	440	420
7	Tremellomycetes	369	314	353
8	Lecanoromycetes	306	306	307
9	Unknown	263	352	360
10	Pezizomycetes	200	195	198
11	Microbotryomycetes	130	129	130
12	Mucoromycetes	126	126	124
13	Orbiliomycetes	123	122	122
14	Glomeromycetes	88	30	84
15	Mortierellomycetes	64	64	64
16	Cystobasidiomycetes	55	46	56
17	Exobasidiomycetes	53	52	53
18	Ustilaginomycetes	43	43	43
19	Pucciniomycetes	41	41	41
20	Agaricostilbomycetes	36	42	42
21	Geoglossomycetes	23	22	22
22	Taphrinomycetes	23	22	23
23	Malasseziomycetes	16	16	16
24	Rhizophydiomycetes	13	13	11
25	Neocallimastigomycetes	10	10	10
26	*Classes with less than 10 sequences	109	104	106

A comparison of the top nine most abundant classes namely, Sordariomycetes, Eurotiomycetes, Leotimycetes,Agaricomycetes, Dothideomycetes, Saccharomycetes, Tremellomycetes, Glomeromycetes, and Unknown (taxa with no resolution) are shown in Figure 11.

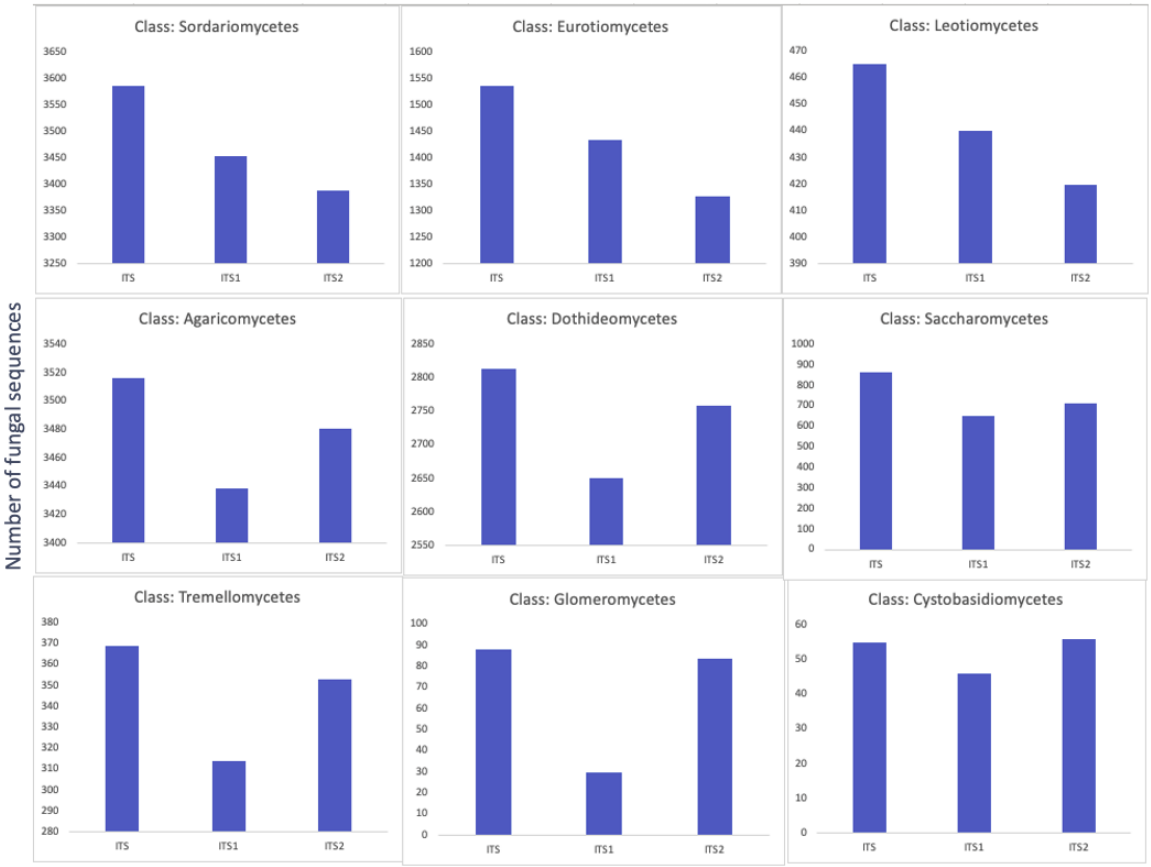

Figure 11. Comparison of the top most abundant taxonomic classes as resolved by each of the three amplicons (ITS, ITS1 and ITS2).

At the class level, the complete ITS performed better than either ITS1 or ITS2 (Fig. 11). In the unknown category, the complete ITS had fewest number of unresolved taxonomic classes as compared to both ITS1 and ITS2, also implying more OTU assignments (Table 5). There however, is one exception to this where ITS underperforms with fewer sequences that were assigned to the taxonomic class: Agaricostilbomycetes as compared to ITS 1 and ITS2 (Table 5).

Some taxonomic classes were better resolved by ITS1 than by ITS 2 (Fig. 11 upper panel). These taxonomic classes included- Sordariomycetes, Eurotiomycetes and Leotiomycetes. Similarly, ITS2 outperformed ITS1 for certain taxonomic classes such as

Agaricomycetes, Dothideomycetes, Saccharomycetes, Tremellomycetes, Glomeromycetes and Cystobasidiomycetes (middle and lower panels (Fig. 11).

Table 6. Summary of overall results for all seven taxonomic levels. The comparison between the number of OTUs for different levels of the taxonomic classification using the three amplicons are shown.

	Original / filtered sequences		Number of operational taxonomic units (OTUs)						
	Source	Sequences	Domain	Phylum	Class	Order	Family	Genus	Species
Observed data	GenBank	15,921	1	8	55	201	647	3,199	15,758
	ITS	14,871	1	15	55	186	574	2,564	6,812
	ITS1	14,118	1	15	56	186	572	2,500	6,779
	ITS2	14,252	1	15	55	186	570	2,489	6,515

V. Discussion

ITSx is a highly effective tool that rapidly merges and trims paired-end FASTQ sequences to the ITS regions of interest for the identification of exact sequence variants (Rivers 2018). The 15,853 sequences were selected post ITSx and used for the AMPtk pipeline to assign operational taxonomic units which were then annotated into various levels of taxonomic classifications and contained complete ITS, ITS1and ITS2 sequences, ensuring fair and equal representation of all three amplicons in the dataset that was processed for AMPtk pipeline. The AMPtk pipeline, with a 97% cut-off for accuracy, provided the most comparable results in terms of accuracy in identifying and annotating the OTUs. Results might vary when different stringencies of cut-off values are used ranging from 97%-100%. Previous studies of LSU and ITS regions showing similar classification accuracies have been reported (Porras et al. 2014). Therefore, it is our understanding that the method chosen to show resolution at the phylum and class level in

Figures 7, 8, and 9, provided comparable classification accuracy at these levels using the 97% cut-off for global alignments to the UNITE database as described in the Materials and Methods section.

A summary of overall results from our analysis at all seven taxonomic levels shows that AMPtk classification agrees fairly well with original annotation at higher taxonomic levels, i.e., class, order and family, but not at lower levels i.e., genus and species (Table 6). Please note that these values are observed data, and no tests for significance were done. A comparison to GenBank data was included to highlight differences in performance of three amplicons, and not as a measure for accuracy.

Sixty-three percent (6,060 OTUs) of the sequences were completely resolved by ITS, ITS1 and ITS2 (Figure 6). We were able to observe a similar study using BLAST where it was shown that most of the sequences obtained for similar ITS fragments could be attributed to fungal organisms (Mello et al. 2011). In this study we established that fungal species can be identified with unique individual markers that are representative of their diversity. Christ et al. (2011) supports this claim and adds that it is due to the high variability and resolution of the unique ITS sequence.

At the phylum level, ITS2 performed extremely well by showing the least unresolved sequences in comparison to ITS1 and the entire ITS region (Table 4, Fig. 10).

Results obtained from this study suggest that future studies aiming to fully resolve Basidiomycota or Glomeromycota, would benefit from using the ITS2 amplicon to understand fungal community structure. The ITS2 region was shown to have a higher

discriminative power to the species level within the Glomeromycota, which corresponded to a larger variability in the ITS2 sequence (Stockinger et al. 2010). Additionally, the ITS2 region is more length conserved than ITS1, making it a more predictable amplicon to work with (Hollingsworth et al. 2011).

The proposed hypothesis that when comparing ITS1 and ITS2 for phylogenetic resolution, an overlapping set of taxa is expected to be identified by both approaches, and certain taxa to be resolved by individual ITS amplicons, was supported by the results according to our venn diagram (Fig. 6). The results were in support of the argument that ITS1 and ITS2 provided specific yet overlapping resolution of the fungal taxa when testing against the fungal community of diverse fungal species belonging to varied taxonomic groups. Another question that remains is if the resolution provided by the complete ITS amplicon is cumulative of the two individual amplicons or not.

It is important to note that though ITS2 is the best choice for fungal resolution at the phylum level, at the taxonomic class level, the entire ITS region is the most efficient choice. We observed the complete ITS outperforming ITS1 and ITS2, without a doubt (Fig. 11). However, using the complete ITS region will always be limited by the read size of the next generation sequencing methods in the metagenomic datasets. Mello et al. (2011) performed a study on ITS1 versus ITS2 pyrosequencing and gave an in-depth comparison of fungal populations in truffle grounds. This significant study used very similar techniques to the ones that were used in our study and was able to attribute many similar conclusions and results. The use of the full ITS region was shown to be a more

time consuming and expensive way to determine fungal diversity in fungal populations (Mello et al. 2011).

It is evident when observing the taxonomic resolution (Figs. 7, 8, 9) that Ascomycota and Basidiomycota were the largest phyla captured in all three amplicons, with Ascomycota being responsible for almost 70% of the entire phyla. This was also established in another study where differences were present in the taxa percentages between ITS's, but each still revealing Ascomycota and Basidiomycota as the dominant fungal phylum (Mello et al. 2011). Other phyla including Glomeromycota, Mortierellomycota, Mucoromycota, and Neocallimastigomycota were under 5% in the resolution graphs using Phyloseq, similar to findings observed in Mello et al. (2011).

At the class level, there were three taxonomic classes within Ascomycota that were resolved the most across all three amplicons (Table 5). Those top, most abundant classes were Sordariomycetes, Agaricomycetes, and Dothideomycetes making up over 75% of the resolution (Fig. 7, 8, 9). A greater percentage of Dothideomycetes, Eurotiomycetes and Sordariomycetes were observed in Mello et al. (2011) Supplemental Figure 2. This suggests that for future studies that aim to capture fungal diversity of Ascomycota or Basidiomycota at the class level, any ITS (ITS1, ITS2 or the entire ITS region) would be an efficient choice because diversity within this class is captured similarly amongst all three amplicons. Kirk et al. (2008) argued in support of Dothideomycetes as a dynamic fungi comprising of the largest, most phylogenetically diverse class within the largest phylum, Ascomycota which is a reason this class was observed as one of the largest classes across all three amplicons. Dothideomycetes

dominated in comparison to other classes like Saccharomycetes and Leotiomycetes. The abundance of unknown sequences in our data provided a limitation for a higher taxa subdivision, similar to Mello et al. (2011).

One of the peculiarities observed in the data set was that there were very few members for each group at the species level. Of the 15,000+ sequences, every species had 2-10 others that were the same. A fungal data set was deliberately taken from the NCBI which represented thousands of species with a few sequences for each, which is the reason the top, most abundant 1,000 OTUs were chosen to be plotted and serve as representatives of the entire data set. Using the top 1,000 OTUs was a preferable approach for data visualization because it prevented over stacking of the bar plots, however it is not the favorable choice when the number of OTUs are very high such as in our case where we had about ~10,000 OTUs. This skews the individual proportions. Given this reason, we performed an in-depth analysis of the data using all the OTUs that were assigned by AMPtk.

The results indicated that a combination of both ITS1 and ITS2 sequences captured the most fungal diversity. A similar study performing a comparison on the performance of ITS1 and ITS2 as barcodes in amplicon-based sequencing yielded comparable results (Mbareche et al. 2020).

V.1. Future prospects

This is a very preliminary analysis that involves using a reference ITS dataset from NCBI as a blind sample and using it to test the effectiveness of the three amplicons (ITS, ITS1 and ITS2) for fungal taxa resolution. Multiple percent identity cut-offs should

be tried as a criteria for homology to the UNITE database for accurate annotations. Additionally, only a single data analysis approach, AMPtk, was used in this study. The next step would be to test multiple pipelines for their effectiveness to assign taxonomic annotation accurately and effectively to the blind sample. These results can then be compared to the original blind dataset to perform a statistical accuracy assessment of how many of those samples in the blind dataset have been correctly matched with their respective and taxonomic annotations. 'Unknowns' in the source data should be addressed as well as any discrepancy due to naming convention differences between original data and UNITE database would need to be explored in more detail.

VI. Conclusions

Programs like FMGP are extremely resourceful for the study of fungal genomics. This program in addition to others like GenBank are helpful but require regular updates to keep up with the evolution of fungi. Expansion and/or improvements on the current DNA based approaches will provide avenues for improvement in the field of using DNA barcodes for mining fungal diversity as well as for maximizing the use of NGS technology for this purpose. Public and free genomics resources were used to investigate the effectiveness of ITS markers as DNA barcodes. While comparing ITS1 and ITS2 and complete ITS for phylogenetic resolution, the hypothesis that an overlapping set of taxa will be identified by both approaches, and certain taxa to be resolved by individual ITSs, was supported by the results. ITS2 was shown to work best for the phyla Basidiomycota and Glomeromycota. The complete ITS outperformed both ITS1 and ITS2 at both the taxonomic levels that was studied in detail.

VII. References

Bazzicalupo, Anna L., Miklós Bálint, and Imke Schmitt. 2013. "Comparison of ITS1 and ITS2 rDNA in 454 sequencing of hyperdiverse fungal communities." *Fungal Ecology* 6, no. 1: 102-109. DOI: 10.1016/j.funeco.2012.09.003

Blaalid, Rakel, Surendra Kumar, R. Henrik Nilsson, Kessy Abarenkov, P. M. Kirk, and H. Kauserud. 2013. "ITS 1 versus ITS 2 as DNA metabarcodes for fungi." *Molecular Ecology Resources* 13, no. 2: 218-224. DOI: 10.1111/1755-0998.12065

Bengtsson-Palme, Johan, Martin Ryberg, Martin Hartmann, Sara Branco, Zheng Wang, Anna Godhe, Pierre De Wit et al. 2013. "Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data." *Methods in Ecology and Evolution* 4, no. 10: 914-919. DOI: 10.1111/2041-210X.12073

Cavalier-Smith, Thomas. 1998. "A revised six-kingdom system of life." *Biological Reviews* 73, no. 3: 203-266.

Christ, Sabina, Tesfaye Wubet, Susanne Theuerl, Nadine Herold, and François Buscot. 2011. "Fungal communities in bulk soil and stone compartments of different forest and soil types as revealed by a barcoding ITS rDNA and a functional laccase encoding gene marker." *Soil Biology and Biochemistry* 43, no. 6: 1292-1299. DOI: 10.1016/j.soilbio.2011.02.022

Cooke, David EL, and James M. Duncan. 1997. "Phylogenetic analysis of Phytophthora species based on ITS1 and ITS2 sequences of the ribosomal RNA gene repeat." *Mycological Research* 101, no. 6: 667-677.

Desprez-Loustau, Marie-Laure, Cecile Robin, Marc Buee, Regis Courtecuisse, Jean Garbaye, Frederic Suffert, Ivan Sache, and David M. Rizzo. 2007. "The fungal dimension of biological invasions." *Trends in Ecology & Evolution 22*, no. 9: 472-480. DOI: 10.1016/j.tree.2007.04.005

Edgar, Robert C. 2010. "Search and clustering orders of magnitude faster than BLAST." *Bioinformatics* 26, no. 19: 2460-2461. DOI: 10.1093/bioinformatics/btq461

Feofilova, E. P. 2001. "The kingdom fungi: heterogeneity of physiological and biochemical properties and relationships with plants, animals, and prokaryotes." *Applied Biochemistry and Microbiology* 37: 124-137.

Gardes, Monique, and Thomas D. Bruns. "ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts." 1993. *Molecular Ecology* 2, no. 2: 113-118. DOI: 10.1111/j.1365-294X.1993.tb00005.x

Gregory, Brian, Nusrat Rahman, Ananth Bommakanti, Md Shamsuzzaman, Mamata Thapa, Alana Lescure, Janice M. Zengel, and Lasse Lindahl. 2019. "The small and large ribosomal subunits depend on each other for stability and accumulation." *Life Science Alliance* 2, no. 2. DOI: 10.26508/lsa.201800150

Grigoriev, Igor V., Daniel Cullen, Stephen B. Goodwin, David Hibbett, Thomas W. Jeffries, Christian P. Kubicek, Cheryl Kuske et al. 2011. "Fueling the future with fungal genomics." *Mycology* 2, no. 3: 192-209. DOI: 10.1080/21501203.2011.584577

Guarro, Josep, Josepa Gené, and Alberto M. Stchigel. 1999. "Developments in fungal taxonomy." *Clinical Microbiology Reviews* 12, no. 3: 454-500. DOI: 10.1128/cmr.12.3.454

Gutiérrez, Lina A., Lina M. Orrego, Giovan F. Gómez, Andrés López, Shirley Luckhart, Jan E. Conn, and Margarita M. Correa. 2010. "A new mtDNA COI gene lineage closely related to Anopheles janconnae of the Albitarsis complex in the Caribbean region of Colombia." *Memorias do Instituto Oswaldo Cruz* 105: 1019-1025. DOI: 10.1590/S0074-02762010000800011

Marco, Jorge D., Paola A. Barroso, Fabricio M. Locatelli, S. Pamela Cajal, Carlos L. Hoyos, M. Cecilia Nevot, Juan J. Lauthier et al. 2015. "Multilocus sequence typing approach for a broader range of species of Leishmania genus: describing parasite diversity in Argentina." *Infection, Genetics and Evolution* 30: 308-317. DOI: 10.1016/j.meegid.2014.12.031

Hollingsworth, Peter M., Sean W. Graham, and Damon P. Little. 2011. "Choosing and using a plant DNA barcode." *PloS One* 6, no. 5: e19254. DOI: 10.1371/journal.pone.0019254

Horisawa, Sakae, Yoh Sakuma, and Shuichi Doi. 2009. "Qualitative and quantitative PCR methods using species-specific primer for detection and identification of wood rot fungi." *Journal of Wood Science* 55: 133-138. DOI: 10.1007/s10086-008-1011-3

Karlep, Liisi, Tõnu Reintamm, and Merike Kelve. 2013. "Intragenomic profiling using multicopy genes: the rDNA internal transcribed spacer sequences of the freshwater sponge Ephydatia fluviatilis." *PLoS One* 8, no. 6: e66601. DOI: 10.1371/journal.pone.0066601

Kirk, Paul., Paul Cannon, David Minter, and Joost Stalpers. 2008. *Ainsworth and Bisby's Dictionary of the Fungi*. Wallingford, UK: CAB International.

Martin, Kendall J., and Paul T. Rygiewicz. 2005. "Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts." *BMC Microbiology* 5, no. 1: 1-11. DOI: 10.1186/1471-2180-5-28

Martin, Francis. "Sequencing the fungal tree of life. 2011. "*New Phytologist* 190:818-821. DOI: 10.1111/j.1469-8137.2011.03688.x

McMurdie, Paul J., and Susan Holmes. 2013. "phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data." *PloS One* 8, no. 4: e61217. DOI: 10.1371/journal.pone.0061217

Mbareche, Hamza, Marc Veillette, Guillaume Bilodeau, and Caroline Duchaine. 2020. "Comparison of the performance of ITS1 and ITS2 as barcodes in amplicon-based sequencing of bioaerosols." *PeerJ* 8: e8523. DOI: 10.7717/peerj.8523

Mello, Antonietta, Chiara Napoli, Claude Murat, Emmanuelle Morin, Giuseppe Marceddu, and Paola Bonfante. 2011. "ITS-1 versus ITS-2 pyrosequencing: a comparison of fungal populations in truffle grounds." *Mycologia* 103, no. 6: 1184-1193. DOI: 10.3852/10-302

Naranjo-Ortiz, Miguel A., and Toni Gabaldón. 2019. "Fungal evolution: major ecological adaptations and evolutionary transitions." *Biological Reviews* 94, no. 4: 1443-1476. DOI: 10.1111/brv.12534

Nolasco, Samuel, and Alejandro Valdez-Mondragón. 2022. "To be or not to be… Integrative taxonomy and species delimitation in the daddy long-legs spiders of the genus Physocyclus (Araneae, Pholcidae) using DNA barcoding and morphology." *ZooKeys* 1135: 93. DOI: 10.3897/zookeys.1135.94628

Ollinger, Nicole, Alexandra Malachova, Michael Sulyok, Lisa Schütz-Kapl, Nicole Wiesinger, Rudolf Krska, and Julian Weghuber. 2022. "Combination of DNA barcoding, targeted metabolite profiling and multispectral imaging to identify mold species and metabolites in sliced bread." *Future Foods* 6: 100196. DOI: 10.1016/j.fufo.2022.100196

Palmer, Jonathan M., Michelle A. Jusino, Mark T. Banik, and Daniel L. Lindner. 2018. "Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data." *PeerJ* 6: e4925. DOI: 10.7717/peerj.4925

Palmer, J. (n.d.). AMPtk documentation - AMPtk 1.3.0 documentation. amptk.readthedocs.io/en/latest/

Porras-Alfaro, Andrea, Kuan-Liang Liu, Cheryl R. Kuske, and Gary Xie. 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, no. 3: 829-840. DOI: 10.1128/AEM.02894-13

Raja, Huzefa A., Andrew N. Miller, Cedric J. Pearce, and Nicholas H. Oberlies. 2017. "Fungal identification using molecular tools: a primer for the natural products research community." *Journal of Natural Products* 80, no. 3: 756-770. DOI: 10.1021/acs.jnatprod.6b01085

Reller, L. Barth, Melvin P. Weinstein, and Cathy A. Petti. 2007. "Detection and identification of microorganisms by gene amplification and sequencing." *Clinical Infectious Diseases* 44, no. 8: 1108-1114. DOI: 10.1086/512818

Rivers, Adam R., Kyle C. Weber, Terrence G. Gardner, Shuang Liu, and Shalamar D. Armstrong. 2018. "ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis." *F1000Research* 7. DOI: 10.12688/f1000research.15704.1

Rognes, Torbjørn, Tomáš Flouri, Ben Nichols, Christopher Quince, and Frédéric Mahé. 2016. "VSEARCH: a versatile open source tool for metagenomics." *PeerJ* 4: e2584. DOI: 10.7717/peerj.2584

Rotmistrovsky, Kirill, Wonhee Jang, and Gregory D. Schuler. 2004. "A web server for performing electronic PCR." *Nucleic Acids Research* 32, no. suppl_2: W108-W112. DOI: 10.1093/nar/gkh450

Roy, Julien, Florent Mazel, Moisés A. Sosa-Hernández, Juan F. Dueñas, Stefan Hempel, Lucie Zinger, and Matthias C. Rillig. 2019. "The relative importance of ecological drivers of arbuscular mycorrhizal fungal distribution varies with taxon phylogenetic resolution." *New Phytologist* 224, no. 2: 936-948. DOI: 10.1111/nph.16148

Schoch, C. L., Pedro W. Crous, Johannes Z. Groenewald, E. W. A. Boehm, Treena I. Burgess, J. De Gruyter, G. Sybren De Hoog et al. 2009. "A class-wide phylogenetic assessment of Dothideomycetes." *Studies in Mycology* 64, no. 1: 1-15. DOI: 10.3114/sim.2009.64.01

Silva, Roberto N. "Perspectives in Genomics The Future of Fungi in 'omics, era." 2016. *Current Genomics* 17, no. 2: 82. DOI: 10.2174/1389202917021603031748 29

Stockinger, Herbert, Manuela Krüger, and Arthur Schüßler. 2010. "DNA barcoding of arbuscular mycorrhizal fungi." *New Phytologist* 187, no. 2: 461-474. DOI: 10.1111/j.1469-8137.2010.03321.x

Stoeckle, Mark, and Paul Hebert. 2008. "Barcode of life." *Scientific American* 299, no. 4: 82-89. http://www.jstor.org/stable/26000849

Tekpinar, Ayten Dizkirici, and Aysenur Kalmer. 2019. "Utility of various molecular markers in fungal identification and phylogeny." *Nova Hedwigia* 109, no. 1-2: 187-224. DOI: 10.1127/nova_hedwigia/2019/0528

Tkacz, Jan S., and Lene Lange. 2004. *Advances in Fungal Biotechnology for Industry, Agriculture, and Medicine.* New York: Kluwer Academic/Plenum Publishers.

Toju, Hirokazu, Akifumi Tanabe, Satoshi Yamamoto, and Hirotoshi Sato. 2012. "High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples." *PloS One* 7, no. 7: e40863. DOI: 10.1371/journal.pone.0040863

Voigt, Kerstin, Timothy Yong, Paul Kirk, André de A. Santiago, Bruce Waldman, Gareth W. Griffith, Minjie Fu et al. 2021. "Early-diverging fungal phyla: taxonomy, species concept, ecology, distribution, anthropogenic impact, and novel phylogenetic proposals." *Fungal Diversity*: 1-40. DOI: 10.1007/s13225-021-00480-y

Vu, Duong, Marizeth Groenewald, Michel De Vries, Thomas Gehrmann, Benjamin Stielow, Ursula Eberhardt, Abdullah Al-Hatmi et al. 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* 92, no. 1: 135-154. DOI: 10.1016/j.simyco.2018.05.001

Wang, Xin-Cun, Chang Liu, Liang Huang, Johan Bengtsson-Palme, Haimei Chen, Jian-Hui Zhang, Dayong Cai, and Jian-Qin Li. 2015. "ITS 1: a DNA barcode better than ITS 2 in eukaryotes" *Molecular Ecology Resources* 15, no. 3: 573-586. DOI: 10.1111/1755-0998.12325

Wang, Xiao-yue, Si-hao Zheng, Yang Liu, and Jian-ping Han. 2016. "ITS2, a Better DNA Barcode than ITS in Identification of Species in Artemisia L." *Chinese Herbal Medicines* 8, no. 4: 352-358. DOI: 10.1016/S1674-6384(16)60062-X

White, Thomas, Thomas Bruns, Steve Lee, and John Taylor. 1990. "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics." *PCR Protocols: A Guide to Methods and Applications* 18, no. 1: 315-322.

Wolf, Matthias, Shilin Chen, Jingyuan Song, Markus Ankenbrand, and Tobias Müller. 2013. "Compensatory base changes in ITS2 secondary structures correlate with the biological species concept despite intragenomic variability in ITS2 sequences–a proof of concept." *PloS One* 8, no. 6: e66726. DOI: 10.1371/journal.pone.0066726

Wu, Bing, Muzammil Hussain, Weiwei Zhang, Marc Stadler, Xingzhong Liu, and Meichun Xiang. 2019. "Current insights into fungal species diversity and perspective on naming the environmental DNA sequences of fungi." *Mycology* 10, no. 3: 127-140. DOI: 10.1080/21501203.2019.1614106

Yasui, Yasuo. 2020. "History of the progressive development of genetic marker systems for common buckwheat." *Breeding Science* 70, no. 1: 13-18. DOI: 10.1270/jsbbs.18123