I consent to loaning my master's thesis for educational purposes when evidence of borrowers is supplied. A borrower is obliged to properly cite all quoted results and conclusions.

Svoluji k zapůjčení své diplomové práce ke studijním účelům a žádám, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně ocitovat.

CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SCIENCE

Study Programme: Biology
Branch of Study: Genetics, Molecular Biology and Virology

Comparison of ITS nrDNA and alternative markers for fungal metabarcoding in environmental samples

Porovnání ITS nrDNA a alternativních markerů pro metabarcoding hub v environmentálních vzorcích

DIPLOMA THESIS

Author: Bc. Tomáš Zelenka

Supervisor: Mgr. Miroslav Kolařík, Ph.D.

Prague, 2015

Declaration:

I hereby declare that I have written this diploma thesis solely by myself and that all sources, references and literature used or excerpted during elaboration of this work are properly cited. The content of this thesis or its major part was not previously used for obtaining of the same or other academic degree.

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne 14.08.2015

....

Acknowledgement

I express deep gratitude to Mirek Kolařík for advising on project design, for factual insights and assistance in all phases of my project. I do thank Petra Havlíčková, Tomáš Větrovský, Martin Kostovčík, Milada Chudíčková, Petr Baldrian and many other co-workers from the Institute of Microbiology AS CR who have shown a true kindness while giving me priceless pieces of advice or helping me to manage such a large quantity of experiments. Next I would like to express a great gratitude to Jan Borovička (Department of Environmental Geology and Geochemistry, Institute of Geology AS CR), Michal Tomšovský (Department of Forest Protection and Wildlife Management at Faculty of Forestry and Wood Technology, Mendel University in Brno) and others who provided samples involved in the study. And last but not least I would like to thank my friends and family, namely: Barča & Barča, Lenka, Zuzka, Darka, Frank and others, who have supported me so many times, gave me valuable comments and helped me find direction in my career.

This study has been running throughout 2012–2015 with the financial support of LD-COST LD13039, 13-06763S, P504/12/0709 (GA ČR) and 360015 (GA UK). Apart from these projects I would like to thank the foundation 'Nadání Josefa, Marie a Zdeňky Hlávkových' for the financial support of my presentation at IMC10 2014 in Bangkok, Thailand and to the COST management for the chance to have an oral contribution at the FA1103 WG Meeting 2014 in Izmir, Turkey.

Abstract

The study of fungal diversity may lead to many fundamental discoveries and conclusions. Molecular genetics, and particularly high throughput sequencing methods using short DNA fragments as barcodes, has recently experienced a boom. The most frequently used marker for fungal research is the partial region of nuclear ribosomal DNA called ITS (Internal Transcribed Spacer). It occurs in the form of tandem repetitions of up to 200 copies. This fact greatly simplifies its amplification from the environment but also introduces some negatives. One of them can be an existence of intragenomic and intraspecific variability which confounds diversity estimates by exaggerating the real number of species. Using alternative low-copy markers can easily prevent these problems. In this study *EF-1α* and *RPB2* protein-coding genes were compared with traditionally used ITS1 and ITS2 markers. An artificial mock community was created by blending genomic DNA of different fungal lineages. The community was sequenced for all markers and the data were processed according to guidelines commonly used in environmental studies. The results show that ITS2 is unequivocally a more suitable marker for environmental studies than other compared markers. The average coefficient of overestimation was deemed to be approximately two for ITS1, ITS2, but also for *RPB2*. *EF-1α* showed largely increased polymorphism within species and therefore this region is not recommended for environmental studies.

Key Words: Fungi, High Throughput Sequencing, Illumina, Environmental Studies, Diversity, Ecology, Molecular Marker, ITS1, ITS2, *RPB2*, *ACT*, *EF-1α*, *MCM7*

Abstrakt

Studium diverzity hub může ve výsledku vést k mnoha významným objevům a závěrům. Molekulární genetika a konkrétně metody masivně paralelního sekvenování se používají ke studiu ekologie a diverzity hub čím dál tím častěji. Využívá se k tomu krátkých úseků DNA označovaných jako *barcode* markery. Nejčastěji používaným markerem je úsek jaderné ribozomální DNA zvaný ITS (*Internal Transcribed Spacer*). Vyskytuje se v genomu ve formě rozsáhlých repetic až 200 kopií, což značně zjednodušuje jeho namnožení z environmentálních vzorků. Zároveň to ale vzbuzuje také určité obavy kvůli výskytu vnitrodruhové a vnitrogenomové variability. Obě tyto variability mohou být zdrojem silného nadhodnocování odhadů diverzity. Použití alternativních, nízko-kopiových markerů, může zmíněný problém částečně vyřešit. V této studii byly porovnány tradičně používané markery ITS1 a ITS2 s protein-kódujícími geny *EF-1α* a *RPB2*. Smícháním genomových DNA druhů z různých fylogenetických skupin bylo vytvořeno *in vitro* umělé společenstvo. To bylo následně sekvenováno pro všechny zmíněné markery a data byla vyhodnocena dle postupů běžně používaných v environmentálních studiích. Výsledky jednoznačně vyzdvihují ITS2 jako nejvhodnější marker pro studium environmentálních vzorků. Průměrný koeficient nadhodnocení lze očekávat kolem dvou pro ITS1, ITS2, ale i pro *RPB2*. *EF-1α* vykázal značnou heterogenitu sekvencí a nelze ho tak pro environmentální studie doporučit.

Klíčová slova: houby, sekvenování nové generace, Illumina, environmentální studie, diverzita, ekologie, molekulární marker, ITS1, ITS2, *RPB2*, *ACT*, *EF-1α*, *MCM7*

CONTENTS

List of Abbreviations
1. Introduction 7
2. Literature Review 9
2.1 Environmental Studies and Metabarcoding 9
2.2 Influence on Diversity Estimates 9
2.3 Different Views of Sequencing 11
2.4 Fungal Molecular Markers 12
2.4.1 Desired vs. Problematic Features 12
2.4.2 Internal Transcribed Spacer 16
2.4.3 Alternative Markers 17
2.5 Rigours of Primer Choice 23
3. Aims of the Thesis 25
4. Material and Methods 26
4.1 Material Generally Used 26
4.2 Selection of Markers for the Study 28
4.3 Samples Preparation, DNA Extraction, Mock Communities 28
4.4 Polymerase Chain Reaction 30
4.4.1 Optimization and Its Hindrances 30
4.4.2 Samples Amplification 31
4.5 Electrophoretic Separation of DNA 34
4.6 Preparation for Sequencing, Creation of Libraries 34
4.6.1 Purification and Blending Scheme 34
4.6.2 Ligation 35
4.7 Data Processing and Bioinformatics 39
4.7.1 'Small' Mocks Processing 39
4.7.2 'Big' Mock Processing 40
5. Results 43
5.1 Pilot Study (Větrovský *et al.* Submitted) 43
5.2 'Big' Mock Analysis 47
5.2.1 Comparative Analysis of Markers 48
5.2.2 Parallel Comparison of Markers 52
5.2.3 Attempt to Quantify 53
6. Discussion 55
6.1 Four Primer Combination Exclusion 55
6.2 Primers over Marker Selection 55
6.3 Comparative Analysis of Markers 56
6.3.1 ITS2 versus ITS1 56
6.3.2 Protein-Coding Markers 57
6.4 Intragenomic Variability 60
6.5 Species and Biomass Quantification 61
7. Conclusions 62
8. References 63
9. Supplementary I
9.1 Supplementary Table 1 I
9.2 Submitted Manuscript XXV

List of Abbreviations

ACT1	Gene for γ-Actin
AFToL	Assembling the Fungal Tree of Life
ALP	Adapter Ligation Plate
ATP6	Gene for the Subunit 6 of ATP synthetase
BLAST	The Basic Local Alignment Search Tool
BM	The ‘Big’ Mock community
BSA	Bovine Serum Albumin
CAP	Clean up ALP Plate
CEP	Clean up End repair Plate
COXI	Gene for the Cytochrome c OXidase I
EF-1α	Gene for the Translation Elongation Factor 1-α
EFL	EF-Like gene
HTS	High Throughput Sequencing methods
IMP	Insert Modification Plate
INSD	The International Nucleotide Sequence Databases
ITS	Internal Transcribed Spacer
ITS1	Internal Transcribed Spacer region 1
ITS2	Internal Transcribed Spacer region 2
Lig	Assemblage for the Ligation reaction
matK	Gene for the Maturase K
MCM7	Gene for the MiniChromosome Maintenance protein
mitLSU	Gene for the Mitochondrial Large SubUnit of rDNA
mitSSU	Gene for the Mitochondrial Large SubUnit of rDNA
NAD	Gene for the NADH Dehydrogenase
NCBI	National Center for Biotechnology Information
nrDNA	Nuclear Ribosomal DNA
nucLSU	Gene for the Nuclear Large SubUnit of nrDNA
nucSSU	Gene for the Nuclear Small SubUnit of nrDNA
NUMT	NUclear MiTochondrial DNA
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
Q30	Phred Quality score 30
rbcL	Gene for the large-chain of RuBisCO enzyme
rDNA	Ribosomal DNA
RPB1	Gene for the Subunit 1 of RNA Polymerase II
RPB2	Gene for the Subunit 2 of RNA Polymerase II
rRNA	Ribosomal RNA
SM	The ‘Small’ Mock community

1. Introduction

The fungal kingdom is a highly diverse and extremely interesting group of organisms. They are ubiquitous living entities of varied size, shape and ecological function. Fungal species interact with a broad range of other organisms – as mutualists they increase nutrient uptake and generally the fitness of many plant species; as dangerous pathogens they can also harm other organisms. They are involved in organic matter degradation but also in production of important substances. There are fungi in soil, living plants, death organic mather of all types, in an insect gut, in a digestive tract of mammals but also in fresh or salt water. Put simply they are well adapted to various conditions and play an irreplaceable ecological role. Besides this they have huge biotechnological potential and therefore people tend to look for new fungal species. There are evolving opinions of how many fungal species actually exist. Some estimates talk of 1.5 up to 5 million species while only about 70,000 are known so far (Blackwell 2011). Nonetheless, recent studies of global fungal diversity speculate that these numbers are overestimated by 1.5- to 2.5-times (Tedersoo *et al.* 2014). Still, so many undiscovered species could hide a real richness of prospective and useful features. A majority of fungi persist in a discreet mycelial phase without any option of cultivation. This limitation had significantly constrained the possibility of understanding fungal relations thereby a rapid and a strong tool for fungal discovery and determination was demanded. Molecular genetics has meaningfully helped us to detect new fungal lineages, reveal true species affiliations and partially untangle the fungal phylogeny. These findings have a practical employment in ecology, medicine, agriculture as well as in discoveries of new bioactive compounds.

The general aim of environmental studies is to get a holistic assessment of fungal communities. Most recent studies of fungal ecology and diversity and many others have been primarily employing methods of molecular biology, especially sequencing by high throughput sequencing methods (HTS). Even though the sequencing helps, it can also be a source of misinterpretation. One of the main variables is a choice of a suitable molecular marker. Moreover, great demands such as universality and proper specificity are put on primers of the potential marker. The partial region of nuclear ribosomal DNA (nrDNA) called ITS (Internal Transcribed Spacer) has recently been the most used fungal barcode marker and has also been declared the 'gold standard' for fungal studies (Schoch *et al.* 2012). It is a multi-copy region and within a genome occurs in variable numbers of copies. This helps to amplify it from the complex environmental samples but it may also

significantly influence the right interpretation of results due to the existence of intragenomic variability. To prevent such problems there are some alternative low-copy markers which may be used.

In studies of highly diverse environments it's very difficult to isolate or even detect some species. This is why sequencing plays an important role. Many papers summarize that only a few percent of species are cultivable but what about the others? An interesting phenomenon is the rare biosphere which denotes less abundant species that are hardly detectable from a small sample. Nonetheless, they could have an important function in the environment (Sogin *et al.* 2006; Zhan & MacIsaac 2014). Many similar studies employ species quantification on the basis of DNA amount, especially using the ITS marker. This information might, however, be strongly biased through the uneven amplification caused by different GC content or by other reasons. Kauserud *et al.* (2008) pointed out that closely related taxa which have a different living strategy also significantly differ in spore size, which is correlated with the genome size (Veselská & Kolařík 2015). Similarly, closely related taxa can differ in the copy numbers of nrDNA, even within the species (Ide *et al.* 2010). This is dependent on the environmental conditions as well as on the particular life strategy; e.g. r-strategists have bigger proteosynthetic apparatus and therefore faster growth while individuals with lower copy number prove to be more sensitive to mutations and thence more adaptive (Kobayashi 2011). It follows how many aspects one should beware of so that the data remain undistorted.

This study investigates the usability of six fungal molecular markers with a total of eight primer combinations in order to evaluate their potential for environmental studies. The literature review summarizes common problems of diversity studies. The impact on result interpretation is described, mainly regarding distinct features of markers and sequencing principles. Then the most frequently used markers in environmental fungal studies are discussed.

During the practical part of the thesis the *in vitro* mock community was created by blending genomic DNA of 463 species (in total 693 items–including species replicates). The choice of strains was focused especially on soil and endophytic fungi as inhabitants of the most studied hyper-diverse substrates. The community was processed in the same way as other diversity studies, i.e. sequenced using HTS and the data processed according traditional guidelines. The obtained sequencing data with regard to the known inputs of this artificial community helped us to unravel properties of tested markers.

2. Literature Review

2.1 Environmental Studies and Metabarcoding

On the planet Earth live several million fungal species. They serve as food, they are producers of important metabolites and other useful compounds, form vital associations with many kinds of organisms, but they are also momentous pathogens (Blackwell 2011). During the last decade molecular identification through DNA has been established as an essential tool for ecological and conservation studies. There are many examples where traditional biotic surveys fail, such as when significant phenotypic plasticity occurs. Molecular techniques might provide reliable results in this situation, plus they are able to reveal fungi abundance and species richness in the environment with dominant but inconspicuous mycelial phase lacking visible fruiting bodies (reviewed by Anderson & Cairney 2004; Seifert 2009). For similar purposes there is a technique called DNA barcoding. This method uses short DNA fragments to decipher species presence in the sample according to the DNA sequences stored in public databases (Hebert *et al.* 2003). Recently the term metabarcoding has become popular in environmental studies where this short DNA region allows the detection of the true spectrum of organisms living in a particular environment (Taberlet *et al.* 2012).

The high throughput sequencing methods present a powerful tool enabling the use of DNA metabarcoding. There have emerged more than 60 fungal studies (Větrovský & Baldrian 2013; Zelenka 2013) using HTS since 2009 when the first fungal study using HTS was published (Öpik *et al.* 2009). A huge amount of fungi, especially symbiotic and parasitic fungi, is nearly uncultivable. From this point of view the introduction of HTS for environmental sample studies was a historic turning point. This is clearly seen from results of Hibbett *et al.* (2011) performed on data from the NCBI GenBank database (http://www.ncbi.nlm.nih.gov). From the total number of 16,969 clusters achieved (defined at 97% similarity level), 37% (6,230 clusters) corresponded to sequences of purely environmental samples.

2.2 Influence on Diversity Estimates

There are several variables having an impact on proper diversity assessment with approximately equal weight. The choice of sequencing method, the best fitting molecular marker, and appropriate primer design together with choice of most suitable primer site

are all very important. It is not so difficult to find a convenient marker for a particular fungal group. However, in diversity studies the broad range of miscellaneous fungal groups with different requirements has to be covered. Even techniques of DNA extraction may significantly influence latter results (Delmont *et al.* 2011). For example, Thakuria *et al.* (2008) point to the three-fold increase in higher quality DNA while using a different DNA extraction set-up versus traditional methods. This may later bias either the community composition estimation or the DNA quantity and consequently ITS copy number estimates (Feinstein *et al.* 2009). However, probably the main and biggest source of bias is the amplification step during PCR in any of its forms. Up to 2,000-fold differences in DNA quantity after amplification have been shown (Pawluczyk *et al.* 2015). Besides a certain randomness of polymerase chain reaction it is likely due to the different content of GC pairs (Suzuki & Giovannoni 1996; Pawluczyk *et al.* 2015) and complicated secondary structures within DNA templates. The choice of DNA polymerase may also have a huge impact. Oliver *et al.* (2015) have shown a proof-reading polymerase that significantly improved their fungal diversity estimates. On the other hand, enzymes with proof-reading activity may even degrade primers and generally cause problems when non-ideal annealing occurs.

Following from rare biosphere importance a complicated question called singletons have emerged. These are sequences that after clustering at a certain similarity level remain as individuals. Some people hold opinions that their removal can deprive us of principal diversity (Unterseher *et al.* 2011), nevertheless a major community recommend the removal of all low abundant sequences even up to ten reads (Brown *et al.* 2015). In relation to the rare biosphere, there are people who tend to quantify species presence. Undoubtedly this could have important ecological conclusions; however, it depends on the way of quantifying. The usual composition of soil (as a typical fungal environment) consists of several predominant species which might even represent more than 50% of their respective community (Baldrian *et al.* 2012). Ecological studies also point to the limitation of DNA-based environmental studies, which when compared with RNA or enzyme-based experiments can show the real significance of some taxa that create the so-called active community (Baldrian *et al.* 2012). Apart from the active community there is nevertheless a vast amount of hyphae fragments, dormant spores or parts of DNA which exist in soil, without any important ecological role. These are recorded by traditional DNA sequencing which may distort the real functional diversity estimation. There are different approaches to diversity studies where not only DNA

(especially nrDNA) but also other bioactive markers such as phospholipid fatty acids or ergosterol content are used for biomass estimation (Baldrian *et al.* 2013). Taken together, attention should be paid to species or biomass quantification while using DNA quantities and particularly ITS copies of nrDNA, as these methods alone will most likely lead to inaccurate results.

2.3 Different Views of Sequencing

Sequencing is a common method designed for the determination of the physical ordering of bases in nucleic acids. HTS methods surpass the Sanger sequencing in cost effectiveness, their overall miniaturization and high throughput of reactions Liu *et al.* (2012). Individual HTS methods differ from each other mainly by the accuracy and diverse length of obtained reads. This is a crucial fact influencing the choice of potential markers for metabarcoding. Pyrosequencing yields sufficiently long reads which is its major advantage especially for environmental studies. However, newer platforms make less mistakes and are generally cheaper, which allows more detailed characterization of fungal communities. Moreover they tend to produce longer reads thus it is probable they will in the near future replace 454 pyrosequencing (Schmidt *et al.* 2013). Reads of a range between 50–300 bp (> 70% bases higher than Phred quality score Q30 for 300 bp) have become available through HTS (http://www.illumina.com; Liu *et al.* 2012; Quail *et al.* 2012). Taking into account the possibility of overlaying reads sequenced from opposite sites (so-called paired end sequencing) it is possible to get reads of approximately 400–500 bp long. This size is already sufficient for barcoding purposes. Therefore, the Illumina platform seems the best current candidate for ecology and diversity studies with its optimal combination of price, read length and low error occurrence.

It is necessary to realize that commonly used direct sequencing (Sanger) results in a consensus sequence of all present copies of the gene. On the other hand, methods based on cloning of a fragment (HTS) produce sequences of individual parts of multi-copy marker regions. It follows there would be a different interpretation of results, i.e. the polymorphism sites are seen mostly as double peaks in chromatograms for direct sequencing while for HTS they are seen as individual sequences. For environmental studies it is hence impossible to distinguish for example between an intragenomic variation and sequences of unknown species. This is partially solved by clustering at the

97% similarity level which has been empirically determined by many studies (Bjorbækmo *et al.* 2010; Kunin *et al.* 2010; Tedersoo *et al.* 2010; Blaalid *et al.* 2013). Unfortunately among some fungal groups the intragenomic variability exceeds 3% which can result in diversity overestimation, or at a minimum, the incorrect interpretation of data.

2.4 Fungal Molecular Markers

The use of DNA barcoding is controversial among taxonomists and systematists (e.g. Will & Rubinoff 2004). Indeed it has some drawbacks. One problem is that the DNA sequence databases contain sequences which are not tied to vouchered specimens. Nonetheless, there are still many positive features which support the use of DNA barcoding (Schander & Willassen 2005). A part of the mitochondrial gene for cytochrome c oxidase I (*COXI*) was established as one of the most appropriate barcoding markers. Subsequently, it was proposed as the main marker for bioidentification of animals (Hebert *et al.* 2003). Its inappropriateness for fungal research will be discussed later. For plant barcoding the two-locus combination of *rbcL* and *matK* genes (Hollingsworth *et al.* 2009) was eventually recommended. In the next chapters some required and parlous properties of commonly used markers will be described, and later detailed characteristics of the most used fungal markers.

2.4.1 Desired vs. Problematic Features

A length range of reads produced by sequencing methods has already been described. How to interconnect this information with the marker choice and why are short fragments resulting from novel approaches actually such a huge problem? Markers can basically be divided into protein-coding and non-coding. This plays a certain role because of different evolutionary constraints. There is an assumption of shifts at the third codon positions, probably through a contribution of the U-turn of the tRNA anticodon loop (Lehmann & Libchaber 2008). No formula can predict with certainty the length of sequence needed to ensure species diagnosis. However, since 45 bp fragments contain 15 third codon positions with a low selection constraint, with the four kinds of bases present in nucleic acids this gives 4^{15} = 1 billion combinations which is sufficient for successful species delimitation. However, on account of the composition bias within some groups and taking into consideration that the majority of positions of closely related species are constant, the length of fragment demanded for reliably distinguishing species is up to

600 bp (Hebert *et al.* 2003). Hence, improving sequencing methods towards longer reads is still essential.

For a requested region there must be an adequate sequence difference among distinct species, known as an interspecific variation. However, in many particular groups an intraspecific (within species) variation also emerges. A sufficient disparity between these two variations, known as a barcode gap, is necessary for a successful species resolution. Unfortunately it is not so easy, as Wiemers & Fiedler (2007) pointed out for the butterfly family Lycaenidae. In this study, divergent phenotype specimens with a very different karyotype occurred sympatrically without any evidence for interbreeding. Their results, however, hinted at a small barcoding gap leading to false results. Cumulative error based on false positives and false negatives for each threshold is shown in Figure 1; p.13. The optimum threshold value is 2.8%, where error is minimized at 18.0%. This threshold represents a level at which sequences are clustered to bypass an impact of sequencing errors and the influence of an insufficient barcoding gap as well (Wiemers & Fiedler 2007). Similarly the issue of a low barcoding gap among different karyotypes in the previous study could arise among fungi. Fungi in environmental samples consist of haploid vs. diploid cells (yeasts), monokaryotic vs. dikaryotic cells (Ascomycota vs. Basidiomycota) or multinucleate cells in Glomeromycota (Simon & Weiss 2008). A graphic visualization of barcoding gap analysis for three fungal molecular markers is illustrated in Figure 2; p.14.

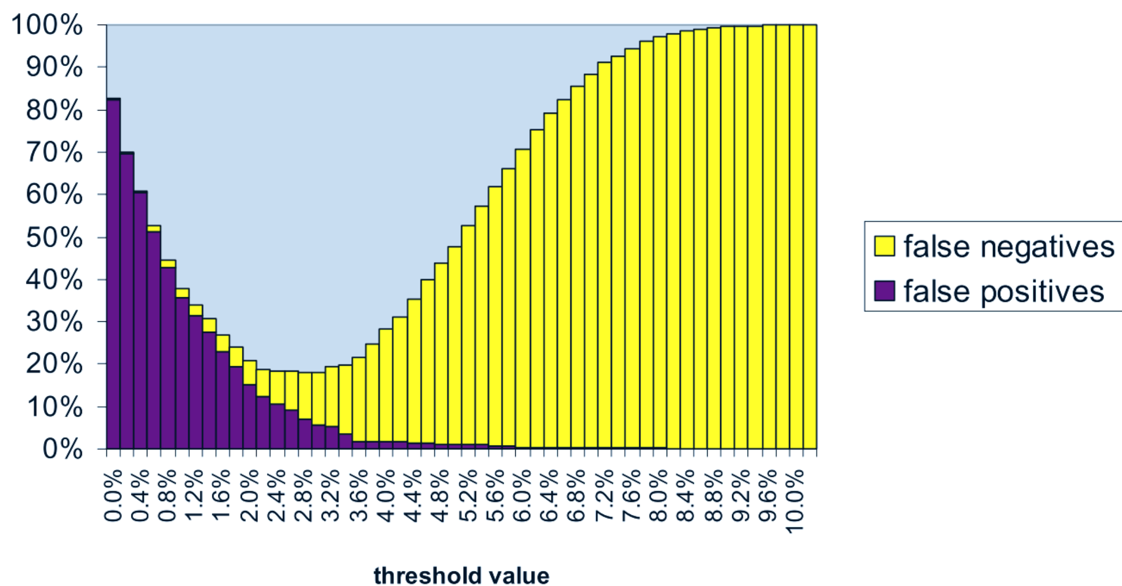

Figure 1. Cumulative error based on false positives plus false negatives for each threshold value in 315 Lycaenidae species including only congeneric comparisons. Taken and adapted from Wiemers & Fiedler (2007).

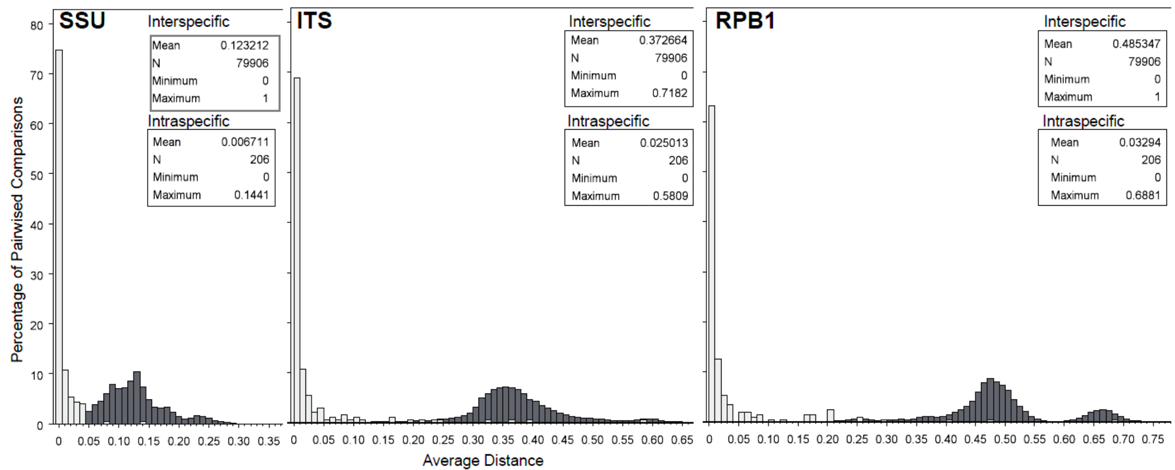

Figure 2. Barcode gap analyses using distance histograms for each marker. Histograms display the intraspecific variation in light gray and the interspecific variation in dark gray. Inserts summarize distance data. It is obvious that the *RPB1* and ITS markers reached substantially distinctive barcoding gap. Taken and adapted from Schoch *et al.* (2012).

Another cause of disorders in environmental studies might be the marker's single-copy or multi-copy nature. Whilst the multi-copy markers are largely derivatives from nuclear and mitochondrial ribosomal DNA (rDNA), the single-copy ones are usually regions of protein-coding genes. Nuclear ribosomal DNA and specifically the ITS region is a crucial marker used for fungal determination and relationship studies among fungal species. This region is also the most widely used gene for assessment of diversity in fungal communities (Schoch *et al.* 2012). Despite its usefulness (discussed later), ITS has many disadvantages (Kiss 2012).

So called deep paralogues, pseudogenes, or the intragenomic variability can all introduce problems into the interpretation of studies. Following from the multi-copy nature of rDNA markers, ITS has a potential to accumulate mutations leading to certain heterogeneity of copies. Regarding the HTS principles this could result in diversity overestimation, data misinterpretation and other issues. For the genus *Laetiporus* it was found that results from sequencing methods based on fragment cloning would overestimate OTU (Operational Taxonomic Unit) diversity using the conventional 97% similarity cut-off more than three times (Lindner & Banik 2011). It can also explain excessive estimates of soil fungal diversity (commonly around thousands of species per gram of soil) based on pyrosequencing results (Jumpponen *et al.* 2010; Hartmann *et al.* 2012; Wubet *et al.* 2012; Monard *et al.* 2013; Orgiazzi *et al.* 2013; reviewed by Zelenka 2013). Data acquired by cloning ITS or using equivalent methods of sequencing are more and more frequent (Jumpponen 2003; Anderson & Cairney 2004; O'Brien *et al.* 2005; Arnold *et al.* 2007; Fierer *et al.* 2007) so the present concerns are justified. The incidence

of the intragenomic variability has been studied in few fungal species so far and any information of greater phylogenetic extent is missing. The ITS region also has great length variability across the sundry fungal groups which could affect mainly PCR reactions (Ihrmark *et al.* 2012). At the same time this region is diverse enough and therefore doesn't allow conducting phylogenetic studies of higher taxonomic resolution.

On the other hand, protein-coding genes are suitable for these purposes with respect to the possibility of employing their amino acid sequences (Větrovský *et al.* Submitted; Glass *et al.* 2013). The most significant trouble of alternative markers could reside in an improper taxonomic classification because of the low representation in public databases. Nevertheless, they are promisingly spreading through an increase of studies using more loci simultaneously (Feau *et al.* 2011). Another difficulty could be the existence of their paralogues or pseudogenes.

This problem might be even larger than the common intragenomic variation occurrence. For instance the deep paralogues – duplicated before the speciation event, occur together among several species and hence they are hardly distinguishable from orthologues (Bailey 2003). Pseudogenes are non-functional genes which may emerge by the gradual amassing of mutations when common reparation mechanisms are relaxed (Feliner & Rosselló 2007) or by gene duplication followed by mutations making the protein coding impossible (Rooney & Ward 2005). Pseudogenes therefore exhibit manifold higher sequence variations and apart from overestimation they can cause serious problems with phylogenetic tree construction (Buckler *et al.* 1997). Mayol & Rosselló (2001) pointed out their importance by comparing two studies performed on oak trees where two distinct phylogenetic trees were created thanks to a probable failure to uncover pseudogenes in one of studies. There are only a few papers studying divergent paralogues among fungi (O'Donnell & Cigelnik 1997; Ko & Jung 2002; Kovacs *et al.* 2011). Some other studies only mention the potential existence of pseudogenes without studying their effect on results (Corradi *et al.* 2007; Rajashekar *et al.* 2007; Rydholm *et al.* 2007; Boon *et al.* 2010; Lafontaine & Dujon 2010; Lindner & Banik 2011). However, the proportion of pseudogenes presumably range between 10–50% in ratio to the functional genes and can't be neglected (Rooney & Ward 2005; Rajashekar *et al.* 2007).

Apart from non-functional paralogues there could be also the functional ones. They are known for the β-tubulin gene (Hubka & Kolarik 2012) and in some particular groups there is an EF-like gene, a paralogue of the commonly used elongation factor 1-α (Keeling & Inagaki 2004; Clouse *et al.* 2013; Kamikawa *et al.* 2013).

2.4.2 Internal Transcribed Spacer

The most common fungal barcode marker is the ITS region of ribosomal DNA. It is split by the 5.8S rRNA gene into the ITS1 and ITS2 regions. Single copies of nrDNA are separated by the non-transcribed intergenic spacer. To date ITS is likely the most appropriate barcode sequence because of the well-defined barcoding gap across a broad range of fungi (Schoch *et al.* 2012), Figure 2; p.14. Nuclear ribosomal DNA occurs in a genome in the form of head-to-tail tandem repetitions in numbers of 20–200 copies (Boyle *et al.* 2005; Raidl *et al.* 2005; Debode *et al.* 2009). This fact substantially facilitates amplification of nrDNA from the environment but also limits its value for quantifying the relative numbers of fungi. Furthermore, ITS is also problematic because of prospective variability among individual copies. Organisms have various ways of dealing with mutations which cause variability. One of them is gene duplication within the scope of birth-and-death evolution as an important homogenization mechanism (Rooney & Ward 2005). Next and perhaps more significant are gene conversion and unequal crossing over (Ganley & Kobayashi 2007) or even deletion of entire nrDNA repetitions during concerted evolution (Ganley & Kobayashi 2011). The number of nrDNA repetitions correlates with the life strategy of fungi. Fungi with low copy number are predisposed to higher DNA damage (Kobayashi 2011), which however leads to an increased adaptability of given organisms. Conversely, fungi with high copy number are able to quickly synthesize proteins and they are rather r-strategists.

Despite all the mechanisms mentioned above, some variability could remain. Intragenomic variation is represented by copy heterogeneity in an individual genome. Simon & Weiss (2008) proved the intragenomic variation 2.2–3.6% among approximately 21% of strains from the group Ascomycota. Moreover Lindner & Banik (2011) showed three-fold overestimation compared with data from Sanger sequencing and also found amplicons with up to 10% variability within the genus *Laetiporus* (Basidiomycota); in this case probably presumed pseudogenes. On the other hand, a complex study aimed at dikaryotic fungi revealed intragenomic variation among only 3–5% of species from a total of 99 species studied (Lindner *et al.* 2013). However, it is still essential to obtain more data, especially from more phylogenetic lineages to definitively exclude effects of intragenomic variation on results of metagenomic studies. It is also important to note that Lindner *et al.* (2013) included only species available in pure cultures. This discriminated large groups of mycorrhizal fungi that are not cultivable but represents a major part of diversity in hyper-diverse forest soils. The intraspecific

variability of ITS among fungi ranges between approximately 0.16–2.85% (Smith *et al.* 2007). According to the International Nucleotide Sequence Database (INSD; http://www.ncbi.nlm.nih.gov) intraspecific variability can be found in about 78% of fungal species with a weighted average value of 2.51%, although its distribution among individual fungal groups is very uneven. Therefore it is not possible to rely only on intraspecific variation influence removal by clustering sequences at 97% similarity level for the OTU definition (Nilsson *et al.* 2008).

The ITS region is usually used as an entire fragment starting in SSU and ending in LSU, but its length is highly variable across fungi (Figure 3; p.18) so the true community composition might be misrepresented because of PCR bias (Ihrmark *et al.* 2012). Nonetheless, the conserved 5.8S region between ITS1 and ITS2 allows placing primers there and splitting of the whole fragment into two parts. This has been increasingly used since the adoption of short-read-length sequencing methods. Although the presence of ITS1 intron in several taxa may cause discarding of these sequences leading to an artificial decrease of OTU richness (Martin & Rygiewicz 2005), there are studies claiming that ITS1 still achieves more faithful results than ITS2 (e.g. Mello *et al.* 2011). Some other results indicate slightly better taxonomic assignments for ITS2 (e.g. Bazzicalupo *et al.* 2013) or similar results for both markers (Blaalid *et al.* 2013). Altogether the usage of both markers simultaneously seems the best option for a huge portion of species (Bazzicalupo *et al.* 2013; Blaalid *et al.* 2013; Monard *et al.* 2013). The research presented here will show the possible existence of paralogues, their abundance and potential impact on study results.

2.4.3 Alternative Markers

Alongside ITS there are also other alternative markers. They differ by their length, taxonomic resolution, abundance within a genome, introns or paralogues present or by universal primer existence. To delineate a range of markers the project AFToL (Assembling the Fungal Tree of Life) can be used as a certain guideline.

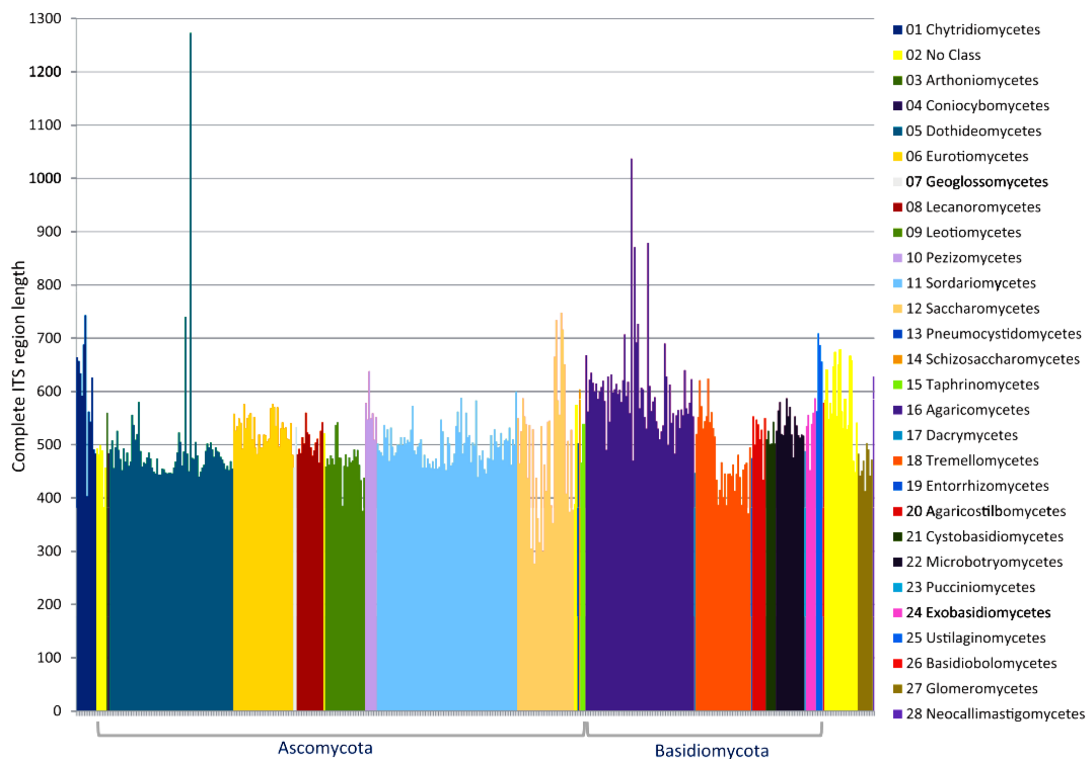

Figure 3. ITS length variation of complete ITS regions in the RefSeq Targeted Loci data set at the class level. Taken and adapted from Schoch *et al.* (2014).

In several studies aimed to uncover the fungal tree of life some multi-loci datasets were used including genes for nuclear and mitochondrial large and small subunits (nucLSU, nucSSU, mitLSU, mitSSU), *5.8S* ribosomal RNA genes, subunits 1 and 2 of RNA polymerase II (*RPB1*, *RPB*2), translation elongation factor 1-α (*EF1-α*), and mitochondrial ATP synthetase (*ATP6*) (Lutzoni *et al.* 2004; James *et al.* 2006; Hibbett *et al.* 2007). Since that time, however, other potentially more suitable markers have been emerging. An important criterion is validity for HTS which eliminates many markers previously used. Hereinafter a brief enumeration of common fungal markers with regard to HTS restrictions will be described.

<u>Small Subunit of nrDNA (nucSSU)</u>

There is no marker used more frequently for bacterial community studies than a partial sequence of 16S (SSU) rDNA (e.g. Rappé & Giovannoni 2003; Eckburg *et al.* 2005). Nonetheless, in the fungal kingdom it has a very low barcoding gap. It evolves relatively slowly so it may be used for studies of early diverging lineages or at least for multi-gene surveys together with ITS or other markers (Schoch *et al.* 2012). It is widely used for phylogeny studies, e.g. for Glomeromycota (Redecker & Raab 2006) or some curious

groups such as Laboulbeniomycetes (Henk *et al.* 2003) and many others (Hibbett *et al.* 2007). Its taxonomic resolution and the number of species recognized by nucSSU in comparison with other markers is, however, very poor and so it is not a good choice for environmental studies (Schoch *et al.* 2012).

Large Subunit of nrDNA (nucLSU)

The partial sequence of nucLSU is a favourite phylogenetic marker. It is very often involved in multi-loci phylogeny and taxonomy studies, especially together with nucSSU (Tang *et al.* 2006; Tretter *et al.* 2013; Voglmayr *et al.* 2013). It is typically used for divergent groups such as arbuscural mycorhizal fungi where ITS usually fails (da Silva *et al.* 2006; Kohout *et al.* 2014). Its D1/D2 region is also frequently used for yeast species characterization (Fell *et al.* 2000; Scorzetti *et al.* 2002). Although, it is very easy to amplify it from the environment due to its multi-copy nature, it has several negatives. Current primer sites are mostly designed for the first 900 bp including D1, D2 and D3 divergent domains as the rest of the gene is invariant even across widely divergent taxa. Most of the primer combinations produce a long fragment improper for HTS. Even though, there are some environmental studies using a partial sequence of nucLSU (Taylor *et al.* 2008; Geml *et al.* 2009); prevailing studies are phylogenetic and taxonomic. nucLSU possesses a low barcoding gap so it is not the best candidate for environmental studies (Schoch *et al.* 2012).

Mitochondrial Large and Small Subunits of rDNA (mitLSU, mitSSU)

Mitochondrial genes generally evolve faster than nuclear genes and therefore mitLSU and mitSSU are suitable for delimitation of lower taxonomic levels (Krak *et al.* 2012). There still remains the advantage of easy amplification through copy multiplication in contrast to low-copy genes.

Mouhamadou *et al.* (2008) showed a great barcoding gap of mitSSU within the genus *Tricholoma*. Their results evinced a very high conservation of sequences within species even among isolates from different geographical origins. Between different species there were substantial sequence differences due to mutations, insertions and deletions allowing indisputable delimitation. However, these insertions were also discovered within the group Polyporales where they caused enormously long sequences (Tomšovský *et al.* 2010). In the same study identical sequences were found for apparently different species (Tomšovský *et al.* 2010). On the other hand it proved a good

discrimination power within *Nomuraea* (Hypocreales) where similar results were reached for nucSSU (Sosa-Gómez *et al.* 2008).

Similarly as nucLSU also mitLSU is used for Glomeromycota studies (Raab *et al.* 2005; Krak *et al.* 2012) and its applicability is obvious also for Ascomycota (Zeng *et al.* 2004) and Basidiomycota (Kretzer & Bruns 1999). A considerable problem may also be an intron presence within some fungal groups (White *et al.* 1990; Raab *et al.* 2005). This can influence PCR success or proper sequence processing. Another difficulty might be the presence of nuclear mitochondrial DNA (NUMTs). These are mitochondrial gene duplicates moved and integrated into the nuclear genome (Lopez *et al.* 1994). Nevertheless, even more crucial is the fact that a reduction of mitochondrial genome in some groups has led to the absolute loss of mitochondrions and hence this kind of genetic information as well. This situation is known to occur for Neocallimastigales and pathogenic Microsporidia (Bullerwell & Lang 2005). These are, however, just minor exceptions and because of their generally good discrimination power they can be used for environmental studies, at least together with another marker.

Additional Mitochondrial Genes

Protein-coding genes are mostly useful for the construction of phylogenetic trees of higher taxonomic resolution. Similarly to mitLSU, mitochondrial protein-coding genes generally show high degrees of heterogeneity and are more practical for studies of lower-level relationships. The most frequently used markers include genes encoding the hydrophobic subunits of the respiratory chain complexes I, III and IV – especially for cytochrome c oxidase (*COX*; several subunites), mitochondrial ATP synthetase (*ATP*; several subunites) and NADH dehydrogenase (*NAD*; several subunits) (Vialle *et al.* 2009).

The *COX1* gene was proposed as a universal barcode marker for animals (Hebert *et al.* 2003). Recently it has been used very frequently across all kingdoms of organisms. Thus the aim of including *COX1* into the fungal barcoding system was more than self-evident. *COX1* proved to have a relatively good discrimination power for fungal species (Seifert *et al.* 2007). At the same time a clear barcoding gap has been visible among fungi and overall *COX1* works well in the study of soil fungi (Molitor *et al.* 2010). The main problem of this region is the presence of introns up to 12 kbp long (Seifert *et al.* 2007; Seifert 2009; Vialle *et al.* 2009). This issue might be avoided by the study of active

community through transcribed RNA where introns are missing (Damon *et al.* 2010). Thus, this marker can be used in environmental studies.

Another option could be a gene for mitochondrial ATP synthetase, probably the frequently used *ATP6* – i.e. for subunit 6. Capabilities of this marker have already been investigated on Boletales (Kretzer & Bruns 1999), nonetheless its usage fails within Ascomycota (Vialle *et al.* 2009).

On the other hand the NADH dehydrogenase subunits worked within Ascomycota (Kouvelis *et al.* 2004; Pantou *et al.* 2006) but the fact that it is missing in several groups of fungi (Bullerwell & Lang 2005) warns against its widespread use.

Gene for *γ*-actin (*ACT1*)

Actin in most fungal groups appears as a single-copy gene per haploid genome (Tarkka *et al.* 2000). Although, it contains introns in some groups, in comparison with other protein-coding genes it is not so problematic (Daniel *et al.* 2001; Lim *et al.* 2004). As it is a protein-coding gene the considerable difference between resolution power of amino acid and nucleotide sequence has been seen (Daniel & Meyer 2003). It is a common advantage of protein-coding genes that the nucleotide sequences possess a significant heterogeneity among each other suitable for species level discrimination. In contrast the conservative amino acid sequence can serve well in higher taxonomic level distinctions. This substantial difference, caused especially by the third position variations, was confirmed through the proper discrimination of yeasts (Daniel *et al.* 2001). In some cases this region worked even better than nucLSU (Daniel & Meyer 2003) and could therefore function as a valuable marker.

Gene for *β*-tubulin

Basically just α-, β- and γ-tubulins genes from the tubulin superfamily are known among fungi (Dutcher 2001). From these the *β*-tubulin is used most frequently; it is even the third most used marker in multi-gene phylogenetic studies (Feau *et al.* 2011). Even though the *β*-tubulin can serve well as a phylogenetic marker in some small groups of fungi (Samson *et al.* 2004; Huang *et al.* 2012; Vela-Corcía *et al.* 2014), it appears less useful in Ascomycota (Hansen *et al.* 2005; Tang *et al.* 2006). No *β*-tubulin sequence has been reported in rust fungi (Ayliffe *et al.* 2001). Simultaneously it possesses a huge number of introns (Ayliffe *et al.* 2001; Shi & Perlin 2001) and one of its main advantages,

its single-copy character, is skewed by the existence of paralogues (Hubka & Kolarik 2012). Therefore it is not a reliable marker for environmental studies.

Translation Elongation Factor 1-*α* (*EF-1α*)

There are several translation elongation factors among fungi. The *EF1α* subunit brings aa-tRNA to the ribosome and has an impact on translational accuracy (Belfield & Tuite 1993). Its ability to recover accurate phylogenetic trees was evaluated in most eukaryotic groups (Roger *et al.* 1999). It also worked well in the study of basal fungal lineages such as Zygomycota (Hoffmann *et al.* 2013), as this was one of the main limits of conventionally used ITS indicated in our study (Větrovský *et al.* Submitted). Conversely, Tanabe *et al.* (2004) warns of poor resolution of phylogenetic relationships within fungi. However, there are still many studies supporting its use (e.g. Brazee *et al.* 2011; Mirhendi *et al.* 2014). A presence of functional paralogue *EF-1α*, EF-like gene (*EFL*), in some particular groups might cause misinterpretation of results (Keeling & Inagaki 2004; Clouse *et al.* 2013; Kamikawa *et al.* 2013). Unfortunately this issue has been poorly investigated among fungi so far. To date only one parallel occurrence of *EF-1α* and *EFL* is known (Henk & Fisher 2012) and in most fungal lineages there probably exists only *EF-1α* or *EFL* (James *et al.* 2006). *EF-1α* is very broadly represented in multi-gene phylogenies (Feau *et al.* 2011) and therefore its applicability should be tested more.

Minichromosome Maintenance Protein (*MCM7*)

This protein serves as a DNA replication licensing factor, i.e. it participates in the pre-replication complex. Although this marker is not as common as the others, many surveys proved its sufficient phylogenetic resolution across diverse fungal groups (Raja *et al.* 2011; Sadowska-Deś *et al.* 2013; Tretter *et al.* 2013). Additionally *MCM7* contained even higher genetic variability than ITS and *RPB1* in another study (Truong *et al.* 2013) and overall it is worth examining *MCM7* as a prospective environmental marker.

Subunits of RNA Polymerase II (*RPB1*, *RPB2*)

Either the largest (*RPB1*) or the second largest (*RPB2*) subunit of RNA polymerase II can provide a very strong support in phylogenetic trees (Hirt *et al.* 1999; Liu *et al.* 1999). Their robust ability to discriminate species and infer phylogenetic relationships has been proven superior to other markers already mentioned. This has been demonstrated across

the broad scale of the fungal kingdom in many studies (e.g. Tanabe *et al.* 2004; Frøslev *et al.* 2005; Hansen *et al.* 2005; Matheny 2005; Tang *et al.* 2006; Schoch *et al.* 2012; Carlson *et al.* 2014). The recent study of Stockinger *et al.* (2014) proved a sufficient barcoding gap and good resolving power for *RPB1* within Glomeromycota and proposed it as a good barcoding marker for this group. Both of them are supposed to be single-copy genes, however, more kinds of sequences of each of them were also reported (Morgenstern *et al.* 2012) as well as a *RPB1* pseudogene and *RPB1* duplication (Matheny 2005). (Schoch *et al.* 2012) pointed out to slightly bigger barcoding gap of *RPB2* over *RPB1* and *MCM7* but at the same time they mentioned their limits in PCR amplification of early-diverging fungi. This is in contrast with our previous results (Větrovský *et al.* Submitted) confirmed also by the results of Tanabe *et al.* (2004). Thus, both of these markers are suggested for environmental studies.

Other Markers

Besides all the markers abovementioned there are many others. Some of them were already examined, some are just timid attempts to enlarge current portfolio. Among those worth mentioning are the intergenic spacer region of rDNA, insertions in the ML5–ML6 region of mitochondrial rDNA, genes for calmodulin, histone H3, chitin synthase I, glyceraldehyde-3-phosphate dehydrogenase, topoisomerase, heat shock proteins, hydrophobin, cellobiohydrolase-C, TwentyS rRNA accumulation protein 1 and many others. Some of them might even possess a certain potential of species delineation. Unfortunately they can only be tested for phylogenetic studies as, they are inappropriate for environmental studies due to their very low representation in sequence databases.

2.5 Rigours of Primer Choice

Choosing a molecular marker is tightly connected with proper primer design. Apart from common variables such as primer length, melting temperature or GC content there are additional constraints in environmental studies. Primers have to be universal enough to efficiently amplify a broad range of distinct organisms. On the other hand they need to be selective to avoid unnecessary amplification of non-target organisms. Study of fungal diversity is obviously intended to study fungi, which means that plant or bacterial DNA itself occupies many of the potential sequencing reads and deprives us of rare species recognition. In most environments just a small group of organisms predominates. In some specific cases an artificial species selection may be used such as in diet composition

studies where excessive concentration of the predator's DNA would preclude the amplification of other minor sequences. Blocking primers may be used against the predator's DNA to enhance sequence specificity for DNA of prey (Vestheim & Jarman 2008). A similar restriction is, however, unfeasible for diversity studies of fungi where dominant species are unknown and additionally vary across many types of studied substrates. Low-copy (protein-coding) markers are sometimes barely amplifiable as pointed out by Schoch *et al.* (2012). Nevertheless, they usually have a better resolving power and from the final pool they are easily detectable by using a sufficient sequencing depth which also allows recovery of rarely occurring sequences (Větrovský *et al.* Submitted; Bellemain *et al.* 2010; Toju *et al.* 2012). Despite the proclaimed universality of ITS primers, they don't cover all fungal lineages *in silico* (Bellemain et al. 2010; Toju et al. 2012). Even though the assumed basidiomycete-specific primer ITS4-B was used, it only amplified a minor proportion of basidiomycete ITS sequences (Bellemain *et al.* 2010). Similar results were obtained in our pilot study where *RPB2* reached a substantially more positive outcome and revealed a broader taxonomic scope (Větrovský *et al.* Submitted).

An ideal primer site is supposed to be a highly conserved sequence flanking some variable region – the marker sequence itself. From this point of view, ITS evinces perfect properties because of its high evolutionary rate of co-occurrence with conservative regions as borders (Begerow *et al.* 2010; Schoch *et al.* 2012). A potential intron existence should be taken into consideration in primer design of protein-coding genes and therefore their placement should target the conservative exon sequence. Some lineages might also be neglected through the codon usage bias within a primer site. The third codon position is very variable, hence primers for protein-coding genes are often degenerated. This fact substantially enhances their universality towards other groups of organisms, even those outside of the fungal kingdom. In spite of the fact that the common ITS primers are considered universal the above-mentioned degeneracy can significantly increase the resolution power of protein-coding markers over the ITS. Regarding new sequencing approaches till now it has been unrealistic to get sequences longer than 300 bp from one side sequencing (http://www.illumina.com). Nevertheless, finding primers for such a short fragment might be really challenging. However, there is a possibility to partially sequence longer fragments as well. The actual contribution a fragment size difference lower than 500 bp could have on PCR or sequencing failure is debatable, which is why I have decided to examine longer markers as well.

3. Aims of the Thesis

The objective of this thesis is to compare ITS nrDNA and other alternative markers for fungal metabarcoding in environmental studies, for diversity and fungal abundance estimates. Our previous results indicate that the *RPB2* marker shows better quantitative representation of taxa than the conventionally used ITS region. At the same time its phylogenetic discrimination power and barcoding gap are superior to ITS (Větrovský *et al.* Submitted). Based on these results other alternative markers were tested. Namely the protein-coding genes for *ACT1*, *MCM7*, *EF-1α*, a different region of *RPB2* and additionally regions ITS1 and ITS2 were chosen. The main goals were:

- The evaluation of taxonomic coverage and resolution across the fungal kingdom
- The comparison of relative distributions within reads and OTU numbers in order to infer quantitative properties of markers
- To reveal potential sources of misinterpretation; i.e. intragenomic variability within multi-copy markers and potential paralogue occurrence within single-copy markers
- To establish the approximate coefficient of overestimation for all markers

4. Material and Methods

This study was based on our previous research. The comparison of *RPB2* and ITS markers was previously performed on environmental samples and evaluated by the artificially created mock community analysis. My contribution to this study was the mock community preparation and its analysis. This ‘Pilot’ Mock community was created by blending genomic DNA of 130 unique and taxonomically distant species. This mixture with known inputs was sequenced using 454 pyrosequencing and analyzed. Detailed methods used in the Pilot study are described in the Submitted Manuscript; p.XXV and Attached on CD (Větrovský *et al.* Submitted).

In this study a ‘Big’ Mock (BM) community was created out of 463 species. However, more than one specimen or strain was selected for some species (693 items in total) in order to infer quantification possibilities for different markers. Conversely to our previous study, Illumina Miseq technology was used for this analysis. Analyzed fungi mostly represented common wood and litter degrading and mycorrhizal fungi as true inhabitants of the environments which are problematic to study. The mock community was amplified with eight primer combinations of six molecular markers. Known inputs of the mock community gave us exact information about primer selectivity and marker suitability which is discussed later.

4.1 Material Generally Used

Solutions:

- TBE electrophoretic buffer - 44mM Tris-HCl, 44mM boric acid, 1mM EDTA, pH=8,0
- TE buffer - 10mM Tris-HCl, 1mM EDTA, pH=7,5

Other Chemicals and Commercial Kits

- ArchivePure™ DNA Yeast & Gram-+ Kit (5 PRIME, Hamburg, Germany)
- MyTaq™ DNA Polymerase (Bioline Reagents Ltd, London, UK)
- PerfectTaq™ Plus DNA Polymerase (5 PRIME, Hamburg, Germany)
- OneTaq® DNA Polymerase (New England Biolabs, Inc., Ipswich, MA, USA)
- DyNAzyme II DNA Polymerase (Thermo Fisher Scientific, Inv., Waltham, MA, USA)
- Omni Klentaq DNA Polymerase (DNA Polymerase Technology, Inc., St. Louis, MO, USA)

- *Pfu* DNA Polymerase (Promega Corporation, Madison, WI, USA)
- SeaKem® LE Agarose (Lonza, Rockland, ME, USA)
- Ethidium bromide 10 mg/ml (Bio-Rad, Hercules, CA, USA)
- 2x DNA Loading Dye (Thermo Fisher Scientific, Inv., Waltham, MA, USA)
- GeneRuler 50 bp DNA Ladder (Thermo Fisher Scientific, Inv., Waltham, MA, USA)
- ZR-96 ChIP DNA Clean & Concentrator™ (Zymo Research Corporation, Irvine, CA, USA)
- Quant–iT™ PicoGreen dsDNA kit (Invitrogen™, Grand Island, NY, USA)
- Qubit® dsDNA HS Assay Kit (Invitrogen™, Grand Island, NY, USA)
- Qubit® dsDNA BR Assay Kit (Invitrogen™, Grand Island, NY, USA)
- MinElute® PCR Purification kit (Qiagen, Hilden, Germany)
- TruSeq DNA PCR-Free LT Sample Prep Kits A and B; Low Sample Protocol (Illumina, Inc., San Diego, CA, USA)
- Agencourt® AMPure® XP (Beckman Coulter, Inc., Pasadena, CA, USA)
- KAPA Library Quantification Kit Illumina® platforms Kapa (Kapa Biosystems, Inc., Wilmington, MA, USA)
- Water Nuclease Free (VWR International Ltd, Lutterworth, England)

Laboratory dishes, microtubes and pipette tips were sterilized in autoclave for 30 minutes (127 °C, 120 kPa). Commercial kits were used according to the manufacturer's protocols and potential divergences are listed in the text.

<u>Equipment</u>

- CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA)
- Qubit® 2.0 fluorometr (Invitrogen™, Grand Island, NY, USA)
- Centrifuge 5804 R (Eppendorf AG, Hamburg, Germany) => Max RCF:
  Fixed-angle rotor 20,913 x g; Swing-bucket rotor 4,500 x g; Plate rotor 2,250 x g
- Mastercycler® Gradient (Eppendorf AG, Hamburg, Germany)
- Mastercycler® Pro S (Eppendorf AG, Hamburg, Germany)
- Equipment for Electrophoresis (Various Providers)
- NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Inv., Waltham, MA, USA)

4.2 Selection of Markers for the Study

Based on the aforementioned theory, mostly markers with good taxonomic discrimination power were chosen for the study. Another criterion for ideal markers was as low as possible paralogue and intron occurrence as well as the existence of primers universal enough to reveal most fungal lineages. Because of the use of Illumina sequencing with limits on fragment size, markers of appropriate length were searched for. Markers chosen for the study are listed in Table 1; p.29, together with primer combinations used for their amplification. Forward primers were, however, tagged in order to distinguish different groups of samples (see Table S3). Markers *RPB2_P* and ITS_P were used in the pilot study and detailed information about the workflow is provided in the Submitted Manuscript; p.XXV and Attached on CD (Větrovský *et al.* Submitted). The region *RPB2_P* represents a part of the *RPB2* gene, while a different *RPB2_B* segment was used in this study to confirm our previous results (Note that these are not official terms. Both are just artificially created in order to distinguish two different regions).

4.3 Samples Preparation, DNA Extraction, Mock Communities

The real usefulness of a marker for environmental studies can be ascertained by comparing known inputs of BM with obtained sequencing data. DNA of most strains subsumed into the BM had already been extracted by and acquired from my colleagues. These analyzed strains came mostly from the Culture Collection of Fungi, Department of Botany, Faculty of Science of Charles University in Prague (strains marked CCF). Other items came from the Institute of Microbiology AS CR, especially from the Culture Collection of Basidiomycetes (marked CCBAS); collections of the Laboratory of Fungal Genetics and Metabolism (marked CFGM), collections of Michal Tomšovský from the Department of Forest Protection and Wildlife Management at the Faculty of Forestry and Wood Technology, Mendel University in Brno (marked CMT); collections of Jan Borovička from the Department of Environmental Geology and Geochemistry, Institute of Geology AS CR (marked CJB); and from the herbarium of the National Museum in Prague (marked PRM). Next, samples were isolated from fresh fruiting bodies collected during mycological excursions (marked CFGM). DNA of these samples was extracted with ArchivePure™ DNA Yeast & Gram-+ Kit according to the manufacturer's instructions. An origin of individual entries is documented in Supplementary Table S1; p.I. and Table S2.

Table 1. List of markers chosen for the study together with used primer combinations. Forward primers were tagged to distinguish fungi which are grouped together. For more detailed information, references and tagged primers see Table S3; Attached on CD. Markers *RPB2_P* and ITS_P were used in the pilot study (Větrovský *et al.* Submitted). The primer combinations which were used for further analysis because of sufficient amplification rate (commented later) are marked in bold.

Marker	Status	Primer Name	Sequence (Direction 5'–> 3')	Length [bp]	Estimated Length of Fragment [bp]	Annel. Temp. [°C]
RPB2_P	Rev.	bRPB2-7R	GAYTGRTTRTGRTCRGGGAAVGG	23	800	66
	For.	bRPB2-6F	TGGGGYATGGTNTGYCCYGC	20		
ITS_P	Rev.	ITS4	TCCTCCGCTTATTGATATGC	20	650	60
	For.	ITS1	TCCGTAGGTGAACCTGCGG	19		
ITS1	**Rev.**	**ITS2**	**GCTGCGTTCTTCATCGATGC**	**20**	**280**	**52**
	For.	**ITS1FI2**	**GAACCWGCGGARGGATCA**	**18**		
ITS2	**Rev.**	**ITS4**	**TCCTCCGCTTATTGATATGC**	**20**	**350**	**52**
	For.	**ITS3**	**GCATCGATGAAGAACGCAGC**	**20**		
EF-1α_A	**Rev.**	**Efgr**	**GCAATGTGGGCRGTRTGRCARTC**	**23**	**380**	**60**
	For.	**EF1-1577F**	**CARGAYGTBTACAAGATYGGTGG**	**23**		
EF-1α_B	Rev.	Efjr	TGYTCNCGRGTYTGNCCRTCYTT	23	450	63
	For.	EF1-526F	GTCGTYGTYATYGGHCAYGT	20		
RPB2_B	**Rev.**	**RPB2-1014R**	**CCRCAIGCYTGICCYTCDGG**	**20**	**390**	**60**
	For.	**RPB2-608F**	**GAYCAYTTYGGIAARAA**	**17**		
ACT1_A	Rev.	Act-3r	TCGGGCAATICITAGGACTITIC	23	250	55
	For.	Act-2	GTCCCIATCTACGAIGG	17		
ACT1_B	Rev.	Act-3r	TCGGGCAATICITAGGACTITIC	23	500	55
	For.	Act-1	TGGGACGATATGGAIAAIATCTGGCA	26		
MCM7	Rev.	Mcm7-1348	GAYTTDGCIACICCIGGRTCWCCCAT	26	560	50
	For.	Mcm7-8af	TGYGGIWSIGARGTITTYCARGA	23		

In total, 463 fungal species were selected to generate the BM. Some strains and specimens of particular species were taken in multiplicities so the total amount of items was 693. All samples were of good PCR quality which was previously tested. Then, DNA was quantified at 260 nm with the NanoDrop 1000 Spectrophotometer and blended together in equimolar proportion. Detailed characteristics of the composition are described in Supplementary Table S1; p.I. and Table S2. Since it would be very difficult to correctly determine all species from such a huge dataset (particularly for markers with low interspecific variability), they were separated into 64 groups to assist in latter identification. These 'Small' Mock communities (SMs) were set up according to phylogenetic relationships of particular fungi. This was ensured by putting together species as phylogenetically distant from one another as possible, at a minimum from

different families (apart from a few exceptions). Within numerous Russulaceae and Strophariaceae families several phylogenetic studies (Shimono *et al.* 2004; Matheny *et al.* 2006) and phylogenetic trees generated by the BLAST (http://blast.ncbi.nlm.nih.gov) platform had to be used for a proper distribution of used species. Therefore strains within one group were supposed to be sequentially distinguishable. In order to differentiate groups from each other they had to be marked. Short oligonucleotide sequences were designed to create eight tagged forward primers for each marker. This tagging was multiplied in a following ligation step where eight labeled Illumina adapters were ligated. This ensured 64 differently marked groups. SMs were created by blending relevant volumes of 10 ng of DNA of each species. In total there were 11 or 12 species included in one SM (Table S2) and all samples were diluted to the final volume of either 110 or 120 μl respectively.

4.4 Polymerase Chain Reaction

4.4.1 Optimization and Its Hindrances

The amplification success of all markers was pre-tested and compositions as well as conditions for PCR reactions were optimized for each marker separately. A tested DNA sample consisted of concentrated DNA of several evolutionarily distinct species, randomly blended. A gradient PCR was run for a range of approximately 10 °C of expected melting temperatures from the reference studies. The most fitting temperatures were used further and are summarized in Table 1; p.29. Touchdown PCR using 0.7 °C steps for the first ten cycles, dropping to ca. 4 °C bellow the melting temperature was performed for all markers with no significant improvement in results. Next, PCR additives Bovine Serum Albumine (BSA; 10 mg/ml) and $MgCl_2$ (25 mM) were tested. After BSA was added, PCRs reached substantially better results. $MgCl_2$ led to apparent non-specificities for some markers but enhanced amplification rate of low-copy markers where 'pure' PCR almost failed. Therefore BSA was added into all reactions whereas $MgCl_2$ was used for selected markers only. MyTaq™ DNA Polymerase; PerfectTaq™ Plus DNA Polymerase; OneTaq® DNA Polymerase; DyNAzyme II DNA Polymerase; Omni Klentaq DNA Polymerase and 4% *Pfu* Polymerase / DyNAzyme DNA Polymerase were examined for problematic markers together with a positive control. The most appropriate enzyme was the PerfectTaq™ Plus DNA Polymerase which was used for most of the following amplifications. Notable is the fact that although the mixture of

proof reading polymerase with another (e.g. 4% *Pfu* Polymerase / DyNAzyme II DNA Polymerase) works well in many studies (Barnes 1994; Arezi *et al.* 2003), a problem may occur when the protein-coding markers are used. The 3'–5' exonuclease-dependent proofreading activity of *Pfu* probably interferes with the primer degeneracy which leads to reaction discontinuance and my study was congruent with this claim.

Eventually, based on the optimization, *ACT1-A*, *ACT1-B, EF-1α_B* and *MCM7* were excluded from further analysis because of their poor amplification success.

4.4.2 Samples Amplification

All PCR reactions were performed in triplicates, irrespective of whether they were 'Big' Mock community or 'Small' Mock community samples. The final volume of each reaction was 25 µl. Each marker was amplified with tagged forward primers and the same reverse primer was used for all samples with regard to the scheme described in Table S4. PCR reaction mixtures were created on ice and the enzyme was added last. Negative controls (no DNA template) were used in every experiment to test for the presence of DNA contamination of reagents and reaction mixtures. Samples were amplified in the Mastercycler® Gradient and Mastercycler® Pro S thermal cyclers.

<u>'Small' Mock Communities Amplificaton</u>

PCR reaction composition for ITS1, ITS2, *EF-1α-A*:

- 16 µl ddH 2 O
- 2.5 µl 10x Buffer for DyNAzyme II DNA Polymerase
- 1.5 µl Bovine Serum Albumine (10 mg/ml)
- 0.5 µl PCR Nucleotide Mix (10 mM; 2.5 mM each)
- 1 µl Forward Primer (Final Concentration 10 µM)
- 1 µl Reverse Primer (Final Concentration 10 µM)
- 2 µl Template DNA
- 0.75 µl DyNAzyme DNA Polymerase (Final Concentration 2 U/µl)

Reaction mixtures didn't contain any additional Mg^{2+} because samples were rich with template and extra Mg^{2+} could introduce more errors and non-specificities which was inferred from previous optimization.

PCR program for ITS1, ITS2, *EF-1α-A*:

- Pre-heating of a Lid – 100 °C
- Initial Denaturation – 94 °C, 5 minutes
- Denaturation – 94 °C, 30 seconds
- Annealing – XX °C, 45 seconds
- Elongation – 72 °C, 45 seconds
- Final Elongation – 72 °C, 7 minutes
- Final Hold – 4 °C, forever

(Denaturation, Annealing and Elongation: 35 x)

Annealing temperature for each marker was used according to Table 1; p.29 – chosen after the optimization.

PCR reaction composition for *RPB2-B*:

- 14 µl ddH 2 O
- 2.5 µl 10x Buffer for PerfectTaq™ Plus DNA Polymerase
- 1.5 µl Bovine Serum Albumine (10 mg/ml)
- 0.5 µl $MgCl_2$ (25 mM)
- 0.5 µl PCR Nucleotide Mix (10 mM; 2.5 mM each)
- 2 µl Forward Primer (Final Concentration 10 µM)
- 2 µl Reverse Primer (Final Concentration 10 µM)
- 2 µl Template DNA
- 0.3 µl PerfectTaq™ Plus DNA Polymerase (Final Concentration 1.5 U/µl)

Reaction mixtures contained additional Mg^{2+} to maximize yield of diverse sequences because of poor DNA content (take into account the low-copy nature of the marker). The enzyme PerfectTaq™ Plus DNA Polymerase was used because during the optimization it had proved substantially better results.

PCR program for *RPB2-B*:

- Pre-heating of a Lid – 100 °C
- Initial Denaturation – 94 °C, 5 minutes
- Denaturation – 94 °C, 45 seconds
- Annealing – XX °C, 30 seconds
- Elongation – 72 °C, 30 seconds
- Final Elongation – 72 °C, 10 minutes
- Final Hold – 4 °C, forever

(Denaturation, Annealing and Elongation: 37 x)

Annealing temperature for the marker was used according to Table 1; p.29 – chosen after the optimization.

<u>'Big' Mock Community Amplification</u>

PCR reaction composition for ITS1, ITS2, *RPB2-B*, *EF-1α-A*:

- 14 µl ddH 2 O
- 2.5 µl 10x Buffer for PerfectTaq™ Plus DNA Polymerase
- 1.5 µl Bovine Serum Albumine (10 mg/ml)
- 0.5 µl $MgCl_2$ (25 mM) (only for *RPB2-B*)
- 0.5 µl PCR Nucleotide Mix (10 mM; 2.5 mM each)
- 2 µl Forward Primer (Final Concentration 10 µM)
- 2 µl Reverse Primer (Final Concentration 10 µM)
- 2 µl Template DNA
- 0.3 µl PerfectTaq™ Plus DNA Polymerase (Final Concentration, 1.5 U/µl)

Reaction mixture of *RPB2-B* contained additional Mg^{2+} to maximize yield of diverse sequences because of poor DNA content (take into account low-copy nature of the marker). For *EF-1α-A*, ITS1 and ITS2 the corresponding volume was substituted with PCR-quality water. In this case samples were too rich with template and extra Mg^{2+} could introduce more errors and non-specificities which were proved during optimization. For all samples PerfectTaq™ Plus DNA Polymerase was used for amplification as it was established the 'top' enzyme during the optimization.

PCR program for ITS1, ITS2, *RPB2-B* and *EF-1α-A*:

- Pre-heating of a Lid – 100 °C
- Initial Denaturation – 94 °C, 5 minutes
- Denaturation – 94 °C, 45 seconds
- Annealing – XX °C, 30 seconds
- Elongation – 72 °C, 30 seconds
- Final Elongation – 72 °C, 10 minutes
- Final Hold – 4 °C, forever

(Denaturation, Annealing and Elongation: 37 x)

Annealing temperature for each marker was used according to Table 1; p.29 – chosen after the optimization.

4.5 Electrophoretic Separation of DNA

Each individual PCR reaction was followed by the product verification by DNA electrophoresis. In all cases 1% TBE gel was used. An adequate amount of agarosis was dissolved by short simmering in TBE buffer. When the temperature dropped to circa 60 °C, solution was stained by ethidium bromide (final concentration 0.4 µg/ml) and poured to plastic electrophoresis tub with a well comb. After the gel solidified, it was placed to the electrophoresis apparatus filled with TBE buffer. The samples of 2 µl volume each were mixed with 1 µl of 2x DNA Loading Dye and loaded to wells. The electrophoresis run with voltage of 6 V/cm for 30 minutes and the product was visualized by UV transluminator.

4.6 Preparation for Sequencing, Creation of Libraries

4.6.1 Purification and Blending Scheme

The PCR products were purified by ZR-96 ChIP DNA Clean & Concentrator and eluted by 25 µl of TE buffer. Next, their concentration was assessed with Quant–iT™ PicoGreen dsDNA kit using the CFX96 Touch™ Real-Time PCR Detection System. All samples were pooled into twelve mixtures intended for the adapters ligation. While blending, not only concentration but the fragment size as well was respected. As a premise it was chosen $9{,}75*10^{10}$ copy numbers corresponding with the amount of 40 ng of medium sized fragment 380 bp which was supposed to be sufficient for the next reactions (according to

the Illumina TruSeq® DNA PCR-Free Sample Preparation Guide). This formula was used to calculate the number of copies:

Number of Copies = (Amount * 6.022×10^{23}) / (Length * 1×10^{9} * 650)

Concentration of these mixtures was verified with Qubit® dsDNA BR Assay Kit for the sample Lig 9 and Qubit® dsDNA HS Assay kit for other samples using Qubit® 2.0 fluorometr. The final concentrations are summarized in Table 2; p.35.

Table 2. Values of concentration for each mixture ready for ligation. In Lig 11 there was higher volume than it is allowed in the ligation protocol (25 ng/μl; correspond to 80 μl). Therefore, MinElute® PCR Purification kit was used and DNA was eluted into smaller volume. Concentration was evaluated with Qubit® dsDNA BR Assay. Additional information about Adapters code for each Lig is provided. Sequences of all adapters are accesible in the Illumina TruSeq® DNA PCR-Free Sample Preparation Guide or in Table S5.

Ligation Number	Lig 1	Lig 2	Lig 3	Lig 4	Lig 5	Lig 6	Lig 7	Lig 8	Lig 9	Lig 10	Lig 11	Lig 12
Concentration [ng/μl]	44	47	54	48	56	56	55	40	86	52	149	42
Volume for 2,000 ng [μl] Needed in Lig Protocol	20	43	37	42	36	36	37	22	24	39	14	48
Adapter Set / Number	A 002	A 004	A 005	A 006	A 007	A 013	A 014	A 015	A 016	A 018	A 019	B 008

4.6.2 Ligation

Ligation steps followed the Low Sample (LS) Protocol of Illumina TruSeq® DNA PCR-Free Sample Preparation Guide with several modifications. The whole procedure contained purification steps, repairing ends, adenylating 3' ends and ligating indexed paired-end adapters.

I. End Repair

This process converts the present overhangs into blunt ends using End Repair Mix 2. The 3' to 5' exonuclease activity of this mix removes the 3' overhangs and the 5' to 3' polymerase activity fills in the 5' overhangs.

1. Lig 1–12 mixtures of volumes according to Table 2; p.35 (containing 2 μg of DNA) were topped up to 50 μl with TE buffer and loaded into the Insert Modification Plate (IMP); one library per well.
2. The thawed End Repair Control and End Repair Mix 2 tubes were centrifuged to 600 x g for 5 seconds.

3. 10 µl of thawed End Repair Control and 40 µl of End Repair Mix 2 were added to each well of the IMP plate with DNA.
4. The entire volume was gently pipetted up and down 10 times to mix thoroughly.
5. The IMP plate was sealed with a Microseal 'B' adhesive seal, put on the pre-programmed thermal cycler and it was run the following program:
   - pre-heating of a lid to 100 °C
   - 30 °C for 30 minutes
   - Hold at 4 °C

II. Clean-up with AMPure Beads (Instead of Size Selection)

1. An original protocol for PCR purification with Agencourt® AMPure® for the 96-well format was followed.
2. The AMPure beads bottle was gently shaken to resuspend any magnetic particles.
3. AMPure beads were added to samples according to equation: volume of AMPure beads = 1.8 x reaction volume, i.e. 100 ul of reaction volume from previous End Repair step was mixed with 180 ul of AMPure beads.
4. The entire volume was gently pipetted up and down 10 times to mix thoroughly.
5. The plate was placed on magnetic stand to separate beads from solution.
6. The cleared solution was aspirated and discarded without disturbing the beads.
7. Beads were washed 2x with 200 ul of 70% ethanol.
8. All ethanol was aspirated and the plate was air-dried completely for 5 minutes at room temperature.
9. DNA was eluted with 20 µl of TE buffer and 15 µl of supernatant was transferred to Adapter Ligation Plate (ALP).

III. Adenylating 3' Ends

1. Thawed A-Tailing Control and A-Tailing Mix tubes were centrifuged to 600 x g for 5 seconds.
2. 2.5 µl of thawed A-Tailing Control and 12.5 µl of thawed A-Tailing Mix were added to each well of the ALP plate.
3. The entire volume was gently pipetted up and down 10 times to mix thoroughly.
4. The APL plate was sealed with a Microseal 'B' adhesive seal, put on the pre-programmed thermal cycler and it was run the following program:
   - pre-heating of a lid to 100 °C

- 37 °C for 30 minutes
- 70 °C for 5 minutes
- 4 °C for 5 minutes
- Hold at 4 °C

5. The ALP plate was removed from the thermal cycler and then proceeded immediately to ligate adapters.

IV. Ligation of Adapters

1. The thawed DNA Adapter, Ligation Control and Stop Ligation Buffer tubes were centrifuged to 600 x g for 5 seconds.
2. 2.5 µl of each thawed Ligation Control, Ligation Mix 2 and DNA Adapter Index (according to Table 2; p.35) were added to each well of the ALP plate.
3. The entire volume was gently pipetted up and down 10 times to mix thoroughly.
4. The ALP plate was sealed with a Microseal 'B' adhesive seal, centrifuged to 280 x g for 1 minute, placed on the pre-programmed thermal cycler and it was run the following program:
   - pre-heating of a lid to 100 °C
   - 30 °C for 30 minutes
   - Hold at 4 °C
5. 5 µl of Stop Ligation Buffer was added to each well of the ALP plate to inactivate the ligation. The entire volume was gently pipetted up and down 10 times to mix thoroughly.

Clean-up of the ALP plate

1. 42.5 µl of well-dispersed Sample Purification Beads were added to each well of the ALP plate. The entire volume was gently pipetted up and down 10 times to mix thoroughly.
2. The ALP plate was incubated at room temperature for 5 minutes and then placed on the magnetic stand at room temperature for 5 minutes.
3. 80 µl of the supernatant from each well in the ALP plate was removed and discarded without disturbing the beads.
4. With the ALP plate remaining on the magnetic stand, 200 µl of freshly prepared 80% EtOH was added to each well without disturbing the beads.

5. The ALP plate was incubated at room temperature for 30 seconds, then all of the supernatant was removed and discarded from each well without disturbing the beads.
6. The steps 4 and 5 were repeated once for a total of two washes.
7. While keeping the ALP plate on the magnetic stand, any remaining EtOH was removed and discarded and the samples were air-dried at room temperature for 5 minutes.
8. While keeping the ALP plate on the magnetic stand, 52.5 µl of Resuspension Buffer was added to each well.
9. The ALP plate was removed from the magnetic stand and the beads in each well were resuspended by repeatedly dispensing the Resuspension Buffer over the bead pellet until it was immersed in the solution, then the entire volume was gently pipetted up and down 10 times to mix thoroughly.
10. The ALP plate was incubated at room temperature for 2 minutes and then placed on the magnetic stand for 5 minutes at room temperature.
11. 50 µl of the clear supernatant was transferred from each well of the ALP plate to the corresponding well of the new 0.3 ml PCR plate labeled with the CAP (Clean up ALP Plate) barcode.
12. 50 µl of mixed Sample Purification Beads were added to each well of the CAP plate for the second clean up. The entire volume was gently pipetted up and down 10 times to mix thoroughly.
13. The CAP plate was incubated at room temperature for 5 minutes and placed on the magnetic stand for 5 minutes at room temperature.
14. 95 µl of the supernatant was removed and discarded from each well of the CAP plate without disturbing the beads.
15. With the CAP plate remaining on the magnetic stand, 200 µl of freshly prepared 80% EtOH was added to each well without disturbing the beads.
16. The CAP plate was incubated at room temperature for 30 seconds, then all of the supernatant was removed and discarded from each well without disturbing the beads.
17. The steps 15 and 16 were repeated once for a total of two 80% EtOH washes.
18. While keeping the CAP plate on the magnetic stand, any remaining EtOH was removed and discarded and the samples were air-dried at room temperature for 5 minutes.
19. While keeping the CAP plate on the magnetic stand, 22.5 µl of Resuspension Buffer was added to each well.

20. The CAP plate was removed from the magnetic stand and the beads in each well were resuspended by repeatedly dispensing the Resuspension Buffer over the bead pellet until it was immersed in the solution and then the entire volume was gently pipetted up and down 10 times to mix thoroughly.
21. The CAP plate was incubated at room temperature for 2 minutes and placed on the magnetic stand for 5 minutes at room temperature.
22. Without disturbing the beads, 20 µl of the clear supernatant was transferred from each well of the CAP plate to the corresponding well of the new 0.3 ml PCR plate.

The DNA content in each sample was quantified using the KAPA Library Quantification Kit Illumina® platforms following the manufacturer's instructions and a final equimolar mixture of all Ligation mixtures was prepared. Sequencing analysis was run on MiSeq Desktop Sequencer using the MiSeq® Reagent Kit v3 (600 cycle) for 2x300 bp long reads in the company GeneTiCA s.r.o., Prague, CZE.

4.7 Data Processing and Bioinformatics

4.7.1 'Small' Mocks Processing

All SMs primarily served the purpose of obtaining reference sequences for latter BM data unraveling – because there is an insufficient reference dataset in public databases for *EF-1α_A* and *RPB2_B*. Paired-end reads from Illumina were joined together using the Fastq-join utility (Aronesty 2013) with minimum overlap of 40 bp and 15% as a maximum difference. Obtained data were filtered and trimmed using the pipeline SEED (Větrovský & Baldrian 2013). Sequences with a quality mean <30; shorter than 50 bp for ITS1, 100 bp for ITS2 and *RPB2_B* and 200 bp for *EF-1α_A*; longer than 500 bp; with a mismatch in the tag; or with ambiguous bases were excluded. Sequences were clustered at the 97% similarity threshold using the USEARCH algorithm (Edgar 2010) to yield OTUs and then singletons were removed. ITS sequences were truncated to entire ITS1 or ITS2 regions using ITS extractor 1.0.8 (Bengtsson-Palme *et al.* 2013) to prevent inaccurate determination and alignment through the extreme 5.8 S similarity among distant taxa. Each OTU was identified according to the most abundant sequence using NCBI BlastN and tBlastX for ITS and protein-coding genes respectively (http://blast.ncbi.nlm.nih.gov). OTUs were assigned to the taxonomic levels using taxonomic information from the NCBI taxonomy server (http://www.ncbi.nlm.nih.gov/taxonomy).

SMs usually contained species assigned to unique families, therefore, the reference sequences were chosen as the most abundant sequence from the OTU cluster corresponding with the proper family for each SM. In several cases, especially among closely related orders with low interspecific variability (some Ascomycota orders), the identity at the order level was used to establish the reference sequence. In a few SMs there were more than one species from the same family (Russulaceae, Strophariaceae, Tricholomataceae) and these were identified using a phylogenetic approach by constructing phylogenetic trees using PhyML 3.0 (Guindon *et al.* 2010) and by mutual comparison. The reference datasets were completed with sequences from Sanger sequencing from my previous studies, from the UNITE database (Kõljalg *et al.* 2013) for marker ITS, PHYMYCO database (Mahé *et al.* 2012) for *EF-1α* and from NCBI GenBank database for all markers (http://www.ncbi.nlm.nih.gov). When it was impossible to find a sequence for the specific species and the genus was unique in the whole BM dataset, a sequence from another species from the same genus was selected. Due to poor representation in databases and relatively low sequencing depth for SMs, from the total of 693 items (including species replicates) in the BM there were eventually only 485 and 445 reference items (including species replicates) for *EF-1α_A* and *RPB2_B*, respectively. A complete reference dataset of 693 items was obtained for both ITS1 and ITS2. From 463 dereplicated species all references for ITS1 and ITS2 were obtained, but only 302 and 258 species for *EF-1α_A* and *RPB2_B* respectively. This information is, however, skewed due to incomplete taxonomic assignment of some items which were identified according to their individual ID number. The quality of reference databases was verified using the BlastN algorithm. Numbers of sequences obtained for all SMs are provided in Table S6 as well as the raw SM data, see the Data Files S1, S2, S3 and S4. The final reference datasets in fasta format are provided as Data Files S5, S6, S7 and S8; see Table S2 for explanation of sequence names.

4.7.2 'Big' Mock Processing

Paired-end reads from Illumina were joined together using the utility Fastq-join (Aronesty 2013) with a minimum overlap of 40 bp and 15% as a maximum difference. Sequences with quality mean lower than 30 were discarded as well as sequences containing one or more errors in the tag or one or more ambiguous bases. The sequences unassigned to the proper marker and non-fungal sequences, i.e. those with a best hit in the NCBI database to the non-fungal taxa, were also removed. ITS sequences were then truncated to entire

ITS1 or ITS2 regions using ITSx extractor 1.0.8 (Bengtsson-Palme *et al.* 2013). Sequences shorter than 200 bp in *EF-1α_A* and *RPB2_B* datasets were removed. ITS1 and ITS2 sequences weren't length filtered with respect to their length variability (seen also in the reference dataset). Datasets were clustered at 97% similarity level against the local reference databases created from SM references (including GenBank sequences and other sources) using the *pick_open_reference_otus.py* including taxonomic affiliation based on RDP Classifier 2.2 (Wang *et al.* 2007) under the QIIME 1.9.1 platform (Caporaso *et al.* 2010). This step also included chimera checking by ChimeraSlayer utility (Haas *et al.* 2011) and both chimeras and singletons were removed. All scripts used during the QIIME processing are provided in Data File S9 and parameters in Data File S10. Clusters directly assigned to the reference were retained. Other clusters identified as apparently not a biological contamination (mostly orders from Basidiomycota) were retained, as well as Ascomycota species corresponding with inputs and determined with more than 99% probability. Some inputs came from fruiting bodies which may be a source of potential environmental contamination such as endophytes of fruiting bodies (typically Ascomycota species). As a precaution therefore, these sequences from Ascomycota orders which are typical environmental contaminants with less than 99% probability, or those which didn't match inputs, were discarded. Incompletely classified Basidiomycota species (which couldn't be an environmental contamination) were retained as potential paralogues or pseudogenes. After filtering environmental contamination, purified datasets were reclustered using the same settings as before. All datasets were resampled to the same sequencing depth of 29,800 according to the lowest number of sequences (ITS2 dataset). Two analyses were conducted as described below.

Parallel Comparison of Markers

In this part, all markers were analyzed separately. To avoid an incomplete database only selected references were used: all references obtained from SMs, and the non-SM references (from Sanger sequencing, public databases, etc.) that were found in BM results for the particular gene. Moreover saturation analyses for selected SMs covering all markers were performed using the Chao1 diversity estimator (Chao 1984). All species from SMs which approached saturation of recovered diversity (when rarefaction curves turn into the plateau phase: 2,000 sequences for ITS1; 1,000 sequences for ITS2 and *RPB2_B* and 800 sequences for *EF-1α_A*) were also included in the reference dataset. Eventually, filtered and purified sequences obtained from the BM were assigned to these

references using RDP Classifier 2.2 (Wang *et al.* 2007) and reads, OTU cluster numbers and observed species were compared with reference inputs.

Comparative Analysis of Markers

Reference datasets for all markers taken from the previous step were overlapped. In total there were 563 overlapping items corresponding to 347 species (75% of inputs) and these are marked in Supplementary Table S1; p.I. Data were analyzed for reads, OTUs and observed species proportional recovery.

5. Results

5.1 Pilot Study (Větrovský *et al.* Submitted)

Hereinafter I will summarize results from mock community assemblage and analysis, as this formed my main contribution to the study. In total, 7,668 sequences remained in the ITS1 dataset after denoising, quality check and chimera removal. The *RPB2_P* dataset was then resampled to the same size. Double clustering (with a 98 and 97% similarity threshold) using the USEARCH algorithm (Edgar 2010) resulted in 275 and 177 OTUs in ITS1 and *RPB2_P* datasets, respectively. Interestingly only 68 species were presented in both datasets, whereas 29 (*RPB2*) and 33 (ITS1) were found in just one of the datasets.

Numbers of obtained reads in both datasets were compared with percentage distribution of real number species in divisions. Even though the basidiomycota-specific primers were used to amplify the *RPB2_P* gene, the *RPB2_P* reads quantity and species representation were in correspondence. On the other hand, the numbers of ITS1 reads amplified with universal primers were highly uneven, strongly biased towards the group Basidiomycota, see Figure 4; p.43.

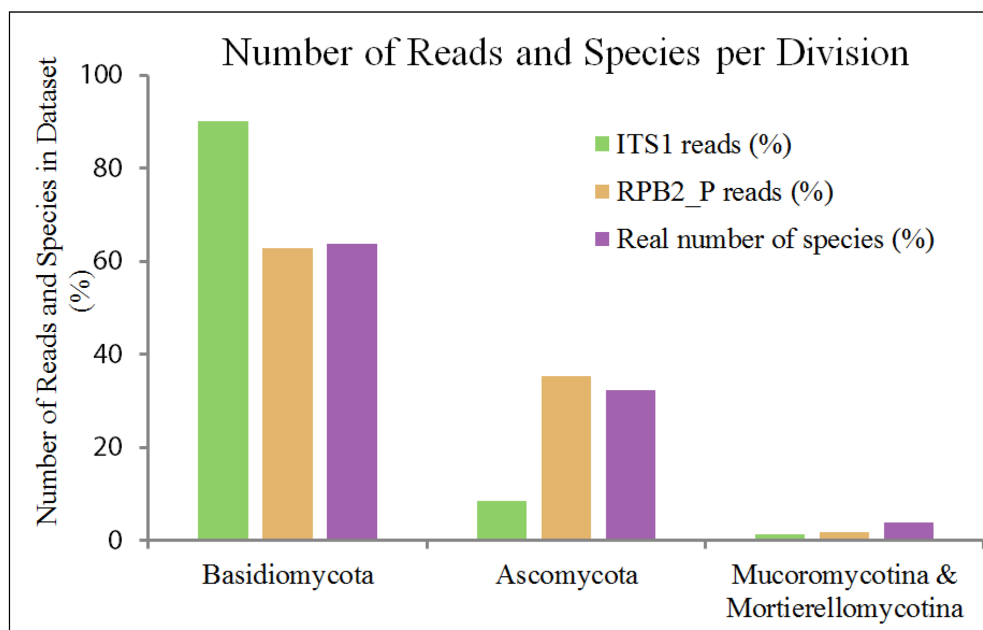

Figure 4. The numbers of obtained reads in both datasets compared with percentage distribution of real numbers of species in divisions. Although the basidiomycete-specific primers were used to amplify the *RPB2_P* gene, there were almost corresponding values between the *RPB2_P* reads quantity and species representation. On the other hand the numbers of ITS1 reads (amplified with universal primers) were highly uneven, preferentially obtained from the group Basidiomycota.

All data were clustered with the commonly used clustering algorithm USEARCH (Edgar 2010). To show implications after using diverse clustering algorithms, two relatively new additional algorithms UPARSE (Edgar 2013) and CROP (Hao *et al.* 2011) were tested. The results showed significant differences among all algorithms. Compared with USEARCH, the new algorithms reached results that were closer to the real species number. Within one dataset there was a difference of more than 100% and depending on the algorithm used in total the numbers of OTUs were overestimated up to 2.7 times (Figure 5; p.44). Nevertheless, one cannot predict the real OTU diversity (regarding all paralogues and variants in a sample) and therefore cannot pronounce any of them the 'best' algorithm.

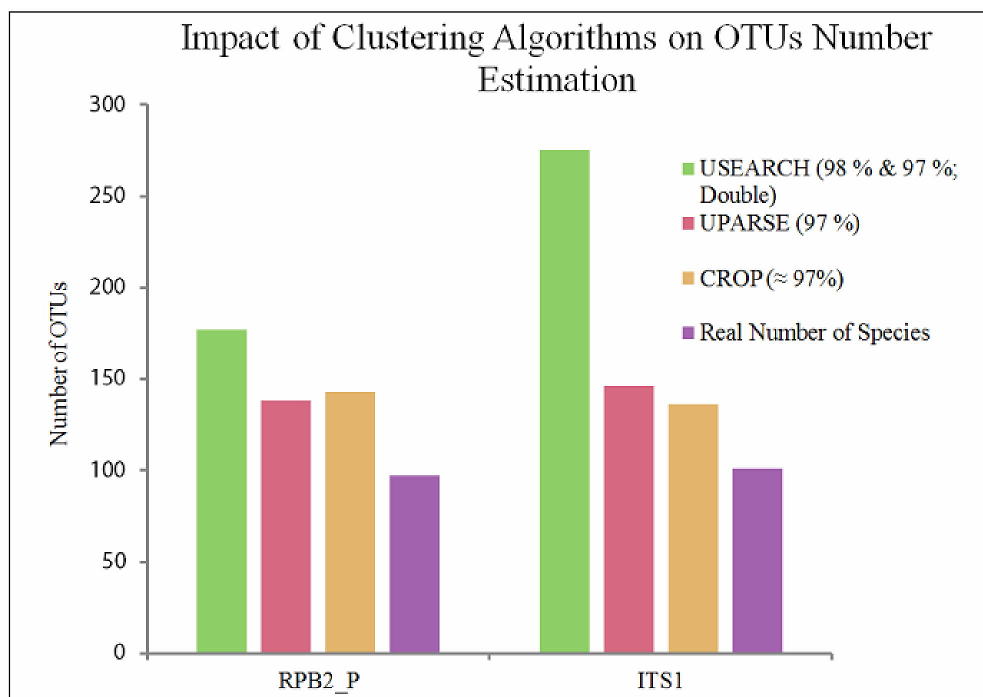

Figure 5. Comparison of three clustering algorithms. The results show a significant overestimation of real species numbers in all cases. However, UPARSE and CROP reached results more closely corresponding to inputs compared with USEARCH. Within the one dataset there was a difference up to 137 % (in ITS1; between USEARCH & CROP).

Fifteen species most abundant in both datasets were randomly resampled at the same sequencing depth (30 sequences) to detect the presence of polymorphisms. ITS1 yielded in 2.53±1.13 OTU per species (Figure 6; p.45) while *RPB2_P* yielded 1.87±0.92 OTU per species (Figure 7 p.46), which is slightly less than ITS1 results ($P = 0.083$).

The distribution of obtained reads almost precisely accorded with an exponential pattern; see Figure 8; p.46. Taken together with the comparison of the 20 most abundant

species, where only seven species were present in both datasets, warns against quantifying on the basis of read numbers. Compared with our results from real environmental samples, single-copy gene *RPB2_P* showed better quantitative and similar qualitative representation of the community composition compared to the conventionally used ITS1 region of nrDNA, and proved applicable as a marker in fungal community studies. However, for better determination of taxa the use of more markers simultaneously is needed, otherwise up to a quarter of real present species could remain hidden. These results were presented at IMC10 2014 in Bangkok, Thailand as a poster presentation, see Figure S3.

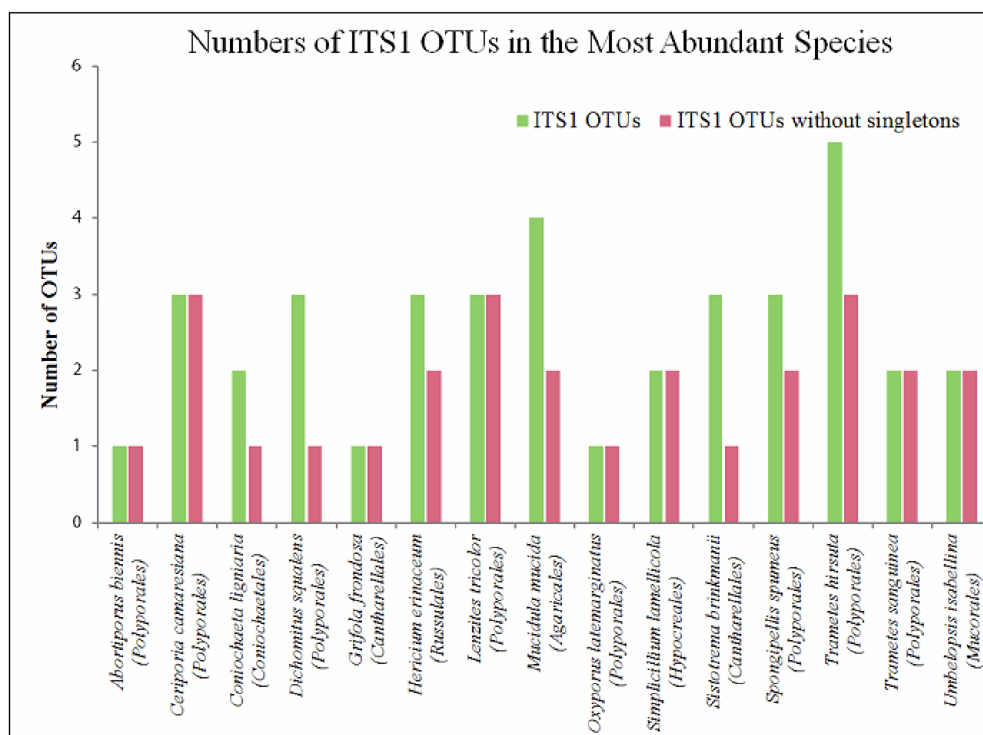

Figure 6. Fifteen species most abundant in both datasets randomly resampled at the same sequencing depth (30 sequences) to show the presence of polymorphisms. In this ITS1 dataset nine species yielded more than one OTU per species (after singletons removal) which is 50 % more than in the *RPB2_P* dataset. This clearly proves the presence of polymorphisms which could lead to distortion of estimated diversity.

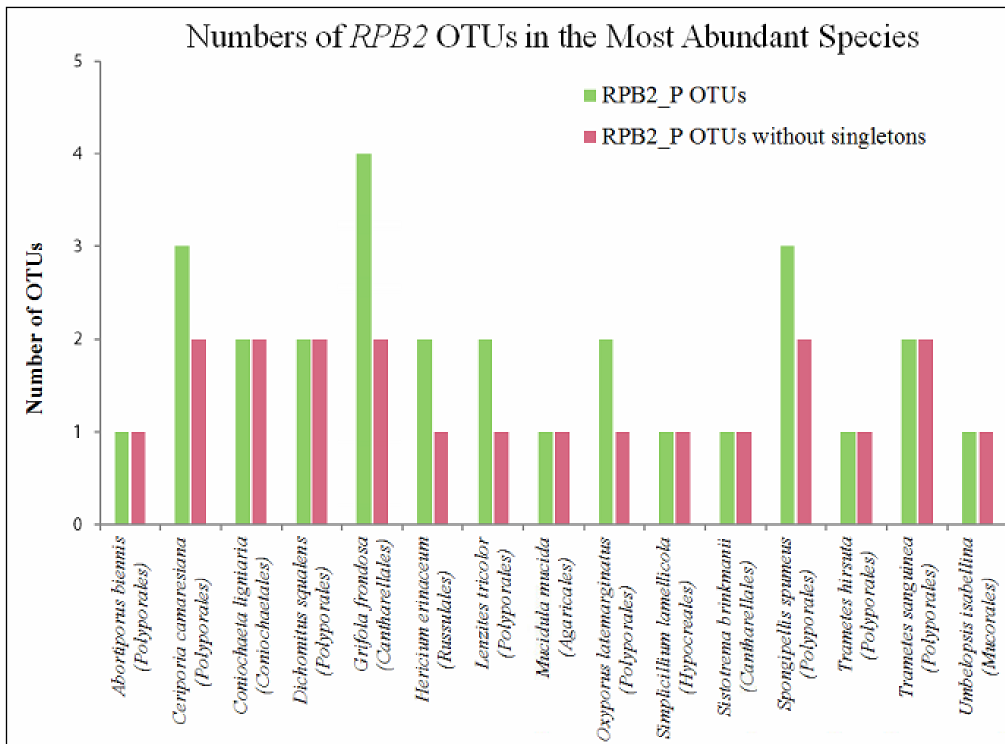

Figure 7. Fifteen species most abundant in both datasets randomly resampled at the same sequencing depth (30 sequences) to show the presence of polymorphisms. In this *RPB2_P* dataset six species yielded more than one OTU per species (after singletons removal) which is 50 % less than in the ITS1 dataset. However, similarly to the ITS1 dataset it still shows how frequent the presence of polymorphisms is, and consequently how much the diversity can be misrepresented.

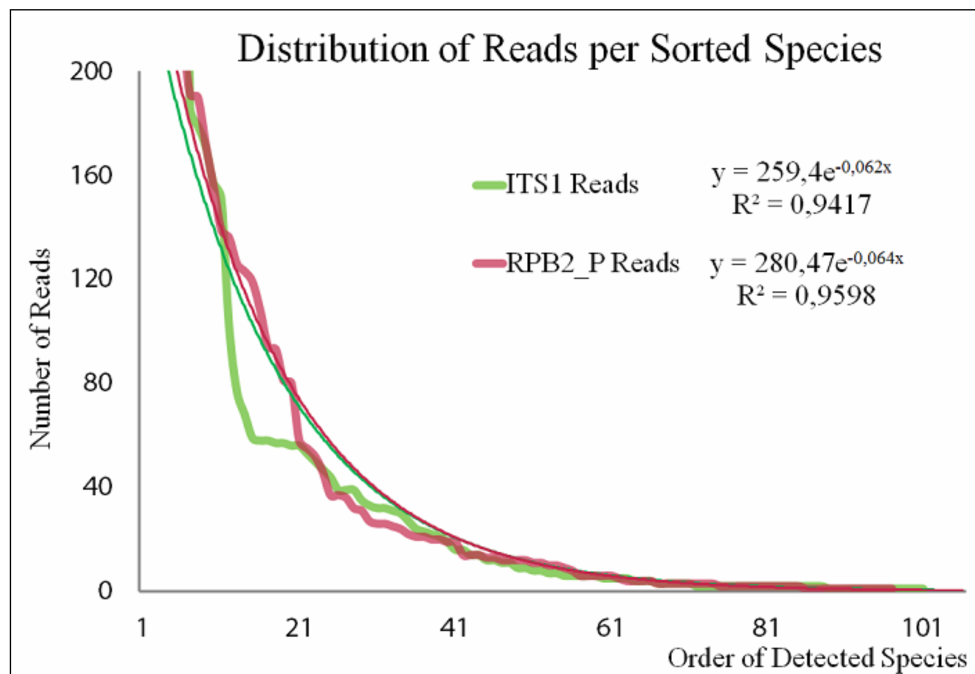

Figure 8. Distribution of obtained reads per detected species shows an almost exponential pattern. These results question the common practice of quantifying species on the basis of read numbers. The most abundant reads of six species are out of chart.

5.2 'Big' Mock Analysis

The BM analysis resulted in 104,239; 31,889; 94,815 and 152,068 filtered and purified sequences for ITS1, ITS2, *EF-1α_A* and *RPB2_B* datasets respectively. Eventually only sequences with order as the lowest taxonomic assignment were used for analyses. This resulted in 29,800 sequences for the ITS2 dataset, and other datasets were then randomly resampled to the same sequencing depth. The detailed sequencing information obtained from data processing is shown in Figure 9; p.47.

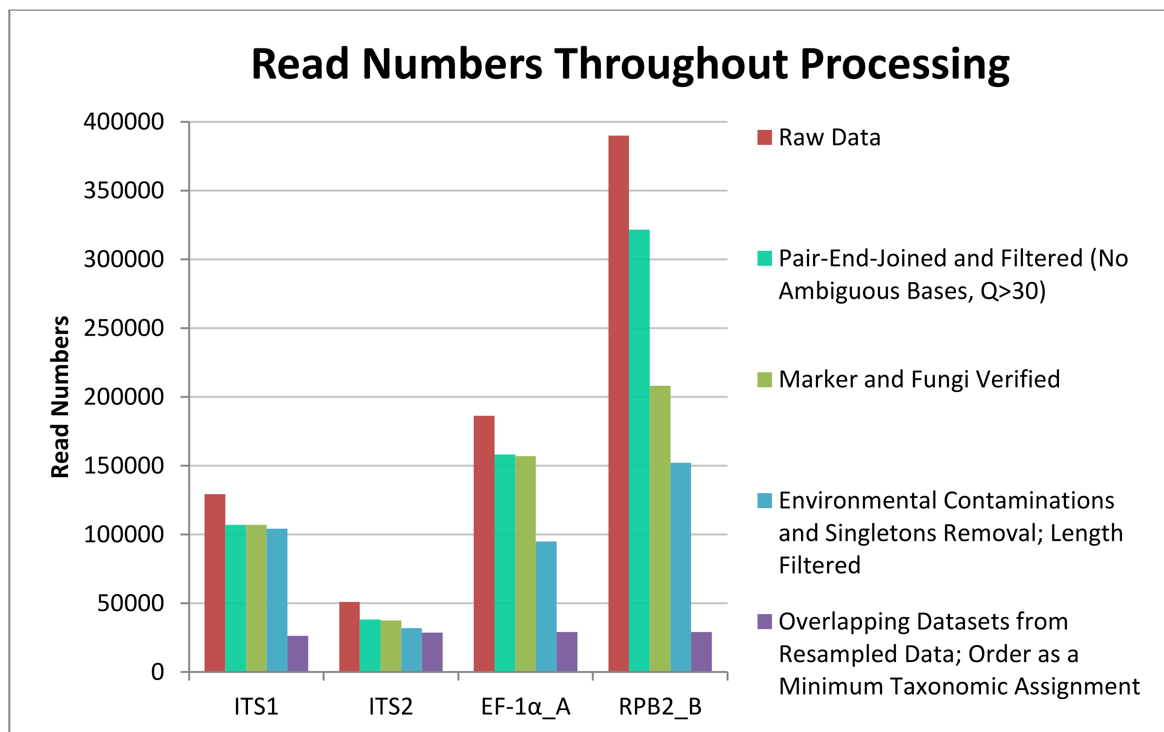

Figure 9. Read numbers throughout data processing.

The sequencing depth was evaluated using the Chao1 diversity estimator (Chao 1984); see Figure 10; p.48. All markers proved to be sufficiently sequenced; i.e. reached saturation. Notice that the OTUs number achieved for the marker *EF-1α_A* greatly exceeds other markers. This kind of high diversity of *EF-1α_A* will be analyzed in the text later. Present diversity within the ITS2 sample was already recovered by ca. 7,000 reads which is interestingly low, compared with other markers (>20,000 reads to reach saturation). It may be caused by a more even distribution of clusters.

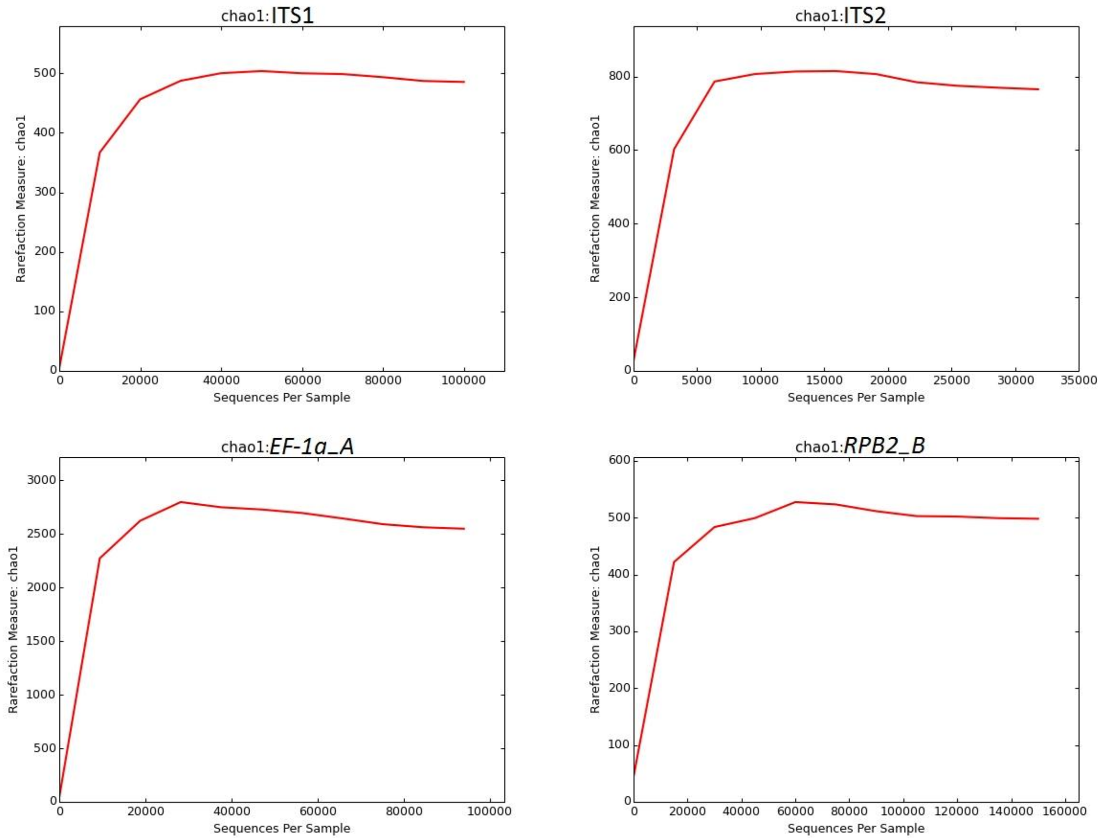

Figure 10. Rarefaction curves computed using Chao1 diversity estimator (Chao 1984). The plots show a number of sequences per sample that are necessary to reveal true diversity within a sample (when the plateau phase occurs).

5.2.1 Comparative Analysis of Markers

During this part of the study different features of markers were compared on the overlapping reference dataset. First, a taxonomic coverage was visualized, see Figure 11; p.49. This graph shows substantial differences among marker abilities to detect OTU presence assigned to at least the order level. Out of 28 orders, nine were not detected at all. However, this may be partially caused by stringent filtering of environmental contaminations. To eliminate the influence of the fact that only the overlapped taxa were studied in this analysis, a parallel comparison of markers for their complete reference datasets was performed in Chapter 5.2.2. A relative comparison of all markers for observed species, OTUs and reads distribution was performed on a dataset including taxa determined at least to the order level, see Figure 12; p.50. At first sight a relatively similar pattern can be seen for all columns. The proportion of sequence references corresponds to their input distribution. Compared to sequence references, species distribution again shows the impossibility of revealing some orders. *RPB2_A* achieved the most even results

when compared with a relative distribution of reads, OTUs and species. Conversely, ITS1 is strongly biased to Agaricales at the reads level. This is due to the enormous recovery of *Entoloma lividum* that reached 87% of all reads in the ITS1 dataset. Despite this fact, ITS1 shows rather comparable results at the OTU level. Nevertheless, this is the first warning against quantification using DNA quantity, particularly the ITS1 region. An overview of correlation coefficients following from comparison of OTUs and observed species, and reads and OTUs for all markers, are shown in Table 3; p.49.

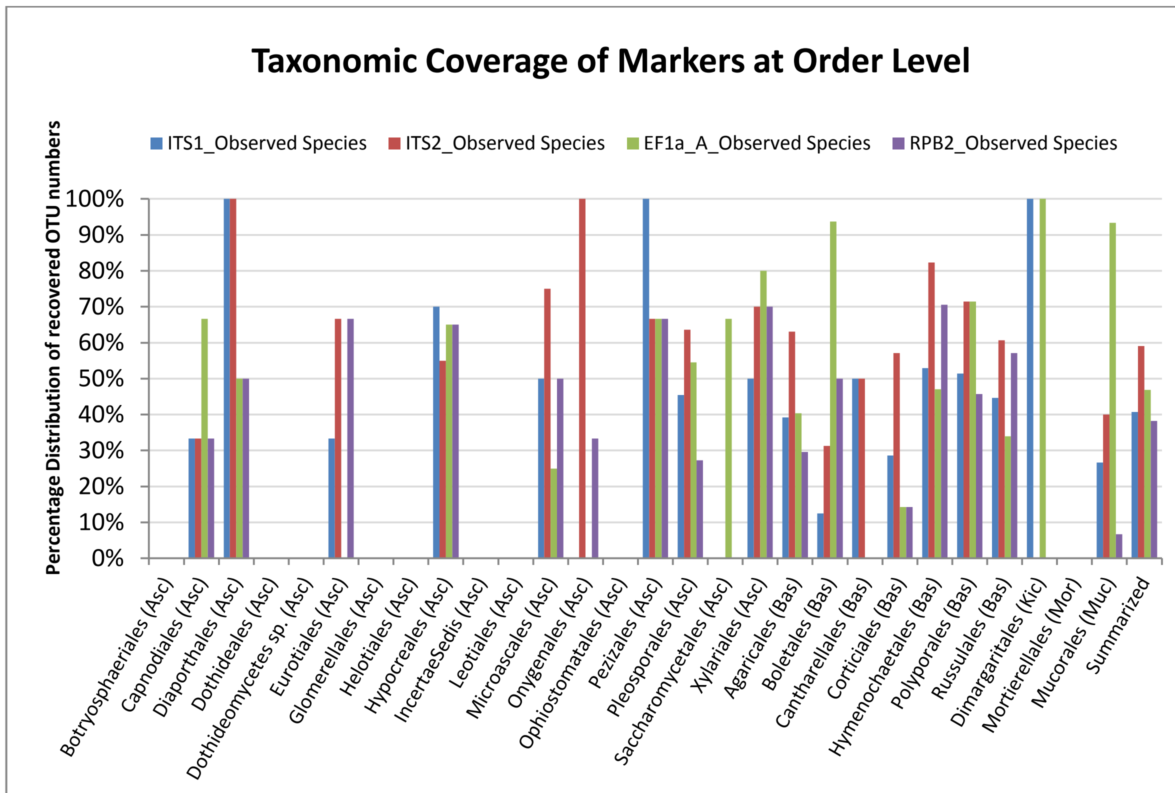

Figure 11. Taxonomic coverage of markers performed on the overlapped reference dataset and filtered to OTU with minimum taxonomic assignment to the order level. In total, nine orders were not detected in analysis.

Table 3. Pearson coefficient of correlation of OTUs numbers to recovered species and read numbers to OTU numbers. Both Parametric and Nonparametric p-values are Bonferroni-corrected; CI= Confidence interval.

Sample ID	Sample ID	Correlation coefficient	Par. P	Nonpar. P	CI (lower)	CI (upper)
ITS1_OTUs	ITS1_Species	0.9946	0.0000	0.004	0.9857	0.9980
ITS2_OTUs	ITS2_Species	0.9980	0.0000	0.004	0.9947	0.9993
EF1a_A_OTUs	EF1a_A_Species	0.9121	0.0000	0.004	0.7818	0.9661
RPB2_OTUs	RPB2_Species	0.9895	0.0000	0.004	0.9722	0.9960
ITS1_Reads	ITS1_OTUs	0.9394	0.0000	0.004	0.8463	0.9768
ITS2_Reads	ITS2_OTUs	0.8221	0.0001	0.004	0.5871	0.9293
EF1a_A_Reads	EF1a_A_OTUs	0.7618	0.0006	0.020	0.4704	0.9034
RPB2_Reads	RPB2_OTUs	0.9754	0.0000	0.004	0.9357	0.9907

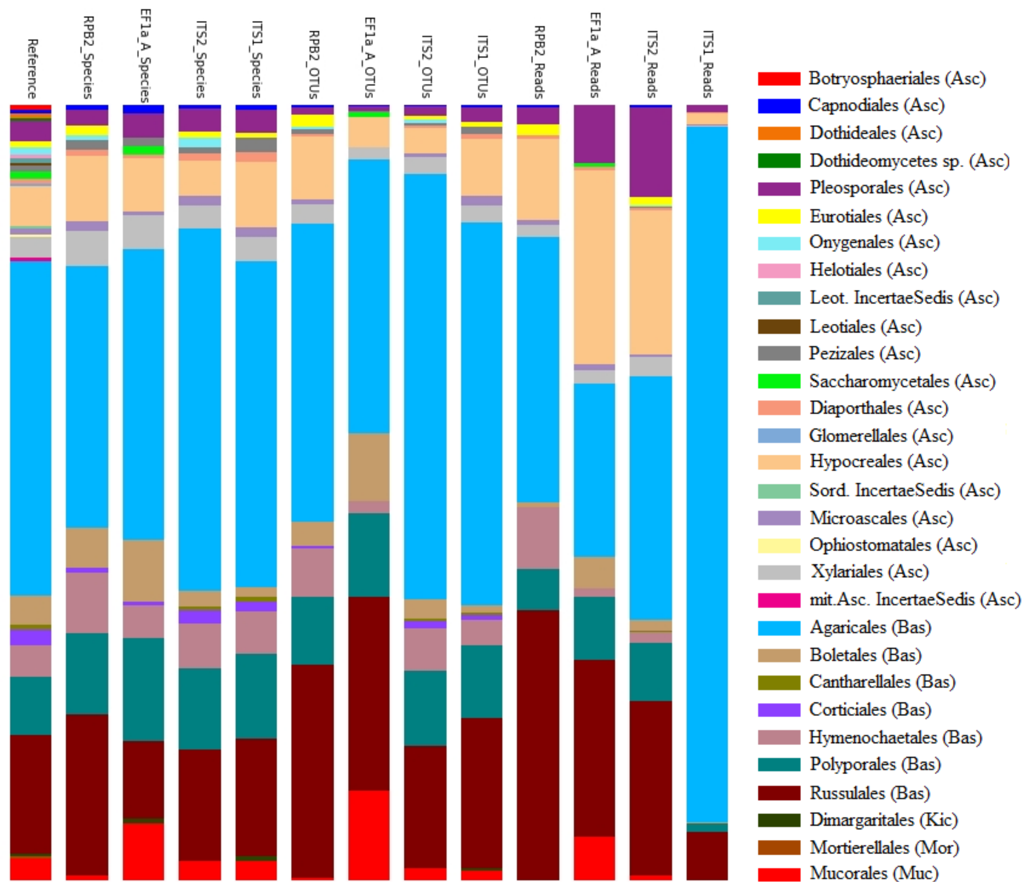

Figure 12. Relative distribution of observed species, OTUs and reads for overlapped reference datasets at the order level. Incertae Sedis groups belong to Leotiomycetes, Sordariomycetes and Mitosporic Ascomycota. Abbreviations Asc., Bas., Kic., Mor. and Muc. denote groups of Ascomycota, Basidiomycota, Kickxellomycotina, Mortierellomycotina and Mucoromycotina respectively. There is an obvious predominance of ITS1 reads numbers for group Agaricales. Nevertheless, OTUs distribution is still more or less similar to other groups. *RPB2_B* show highly correlating distributions within all reads, OTUs, observed species and references.

The presence of intragenomic or intraspecific variability can be found within several taxa. The information about overestimating may be provided by the ratio of obtained OTUs to revealed species, see Figure 13; p.51. Among Ascomycota there are very low ratios of OTUs to species which correspond with their low variability known from literature. Still, orders Eurotiales and Hypocreales show certain heterogeneity by possibly two-fold overestimation. Unfortunately, some Ascomycota paralogues may be missing because of stringent filtering against environmental contaminations. Nevertheless, the study was particularly aimed at the mycorrhizal species which could not be affected by this step. A different situation was apparent within Basidiomycota where most orders showed an increased heterogeneity. The overestimation is extremely high, within several orders for *EF-1α_A*. The summarized ratio value of all obtained OTUs to all observed species overestimates the expected species numbers by 7.8 times. Its average value for all orders

achieved 4.8 ± 4.99. On the other hand, values for other markers were substantially lower, oscillating around the value two. Overall values were 1.9 for ITS1, 2.7 for ITS2 and 2.1 for *RPB2*. This information is, however, incomplete as for many groups there were underrepresented numbers of sequences which preclude establishing a total coefficient of overestimating. More detailed information is provided in supplementary, Figure S1 and S2 (Attached on CD), showing the same ratio at family and species level.

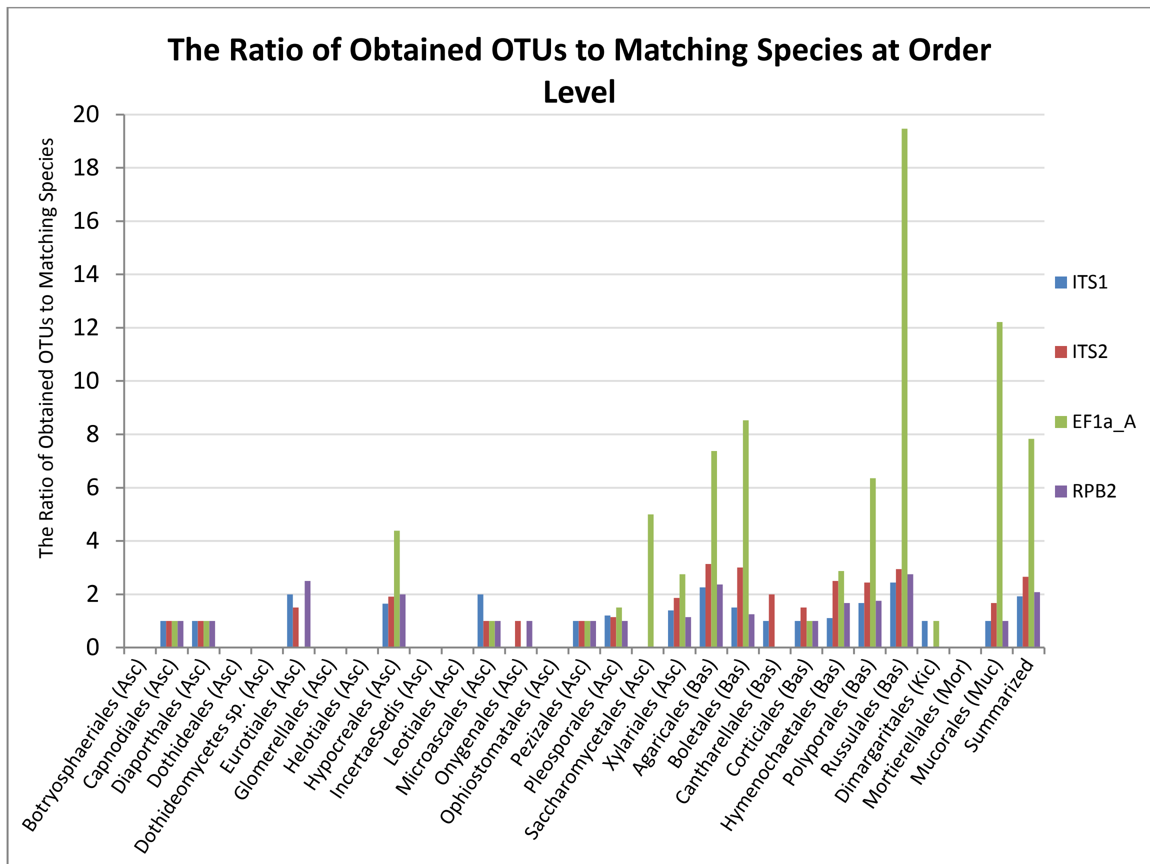

Figure 13. The ratio of obtained OTUs to matching species, calculated at the order level. There are obvious increased numbers of OTU, mainly within Basidiomycota (the right side of the chart), while Ascomycota with generally low variability are not as overestimated. The most striking values belong to *EF-1α_A* which will be discussed later.

From an overestimating point of view, the most noteworthy families are Bondarzewiaceae, Schizophyllaceae, Serpulaceae for all markers and Omphalotaceae, especially for *RPB2_B*. In addition for Entolomataceae, the species *Entoloma lividum* reached an exceptionally high number of reads for ITS1. Nevertheless, the ITS2 dataset with a lower number of reads manifests a more than 15-fold overestimation of species number, similar to ITS1. Only three species of *Entoloma* without any duplicates were present within inputs which point to their likely high intragenomic variability. Other

genera with increased ratios were *Heterobasidion, Cordyceps, Laetiporus, Mensularia, Polyporus, Russula, Schizophyllum* and *Melanoleuca*. These were, however, present in the input dataset in more replicates and thus the true source of misrepresentation may be shared with intraspecific variability. A notable fact is that *EF-1α_A* was highly overestimated mainly in species belonging to *Russula* which will be discussed later.

The results from datasets filtered to species as a minimum taxonomic level are shown as a Venn diagram, Figure 14; p.52. From a total of 274 species (without species multiplication) there were only 44 (16%) species present in all datasets. On the other hand, 4%, 9%, 8% and 6% were detected by only ITS1, ITS2, *EF-1α_A* or *RPB2_B* marker, respectively. From this point of view ITS2 reached the best results as it revealed 75% of species from a total of 274 species covered. This is followed by 56% for *EF-1α_A*, 53% for ITS1 and 50% for *RPB2_B*.

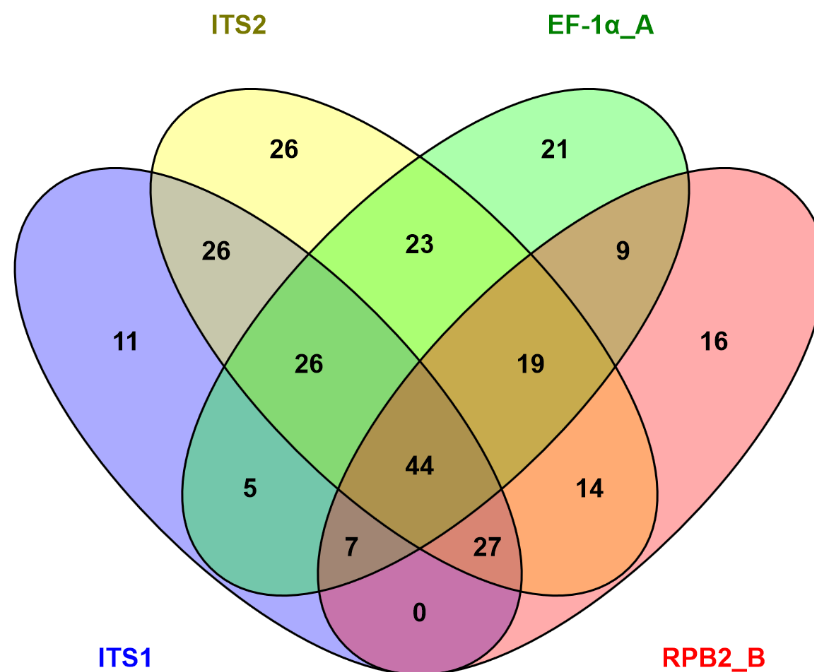

Figure 14. A Venn diagram shows overlaps at the species level for each marker.

5.2.2 Parallel Comparison of Markers

This part of the study focused on taxonomic coverage evaluation. Data were mined from broader reference datasets therefore one can expect a more realistic estimation of their resolution power. Nonetheless, compared to the overlapped reference dataset any larger difference was not evident. The most powerful marker was ITS2 which revealed 33 additional species and had a total coverage of 55% at the species level. ITS1 reached 21

additional species with a total coverage of 40%. Protein-coding genes were only slightly worse; *EF-1α_A* uncovered 13 additional species with coverage of 39% and *RPB2_B* found only five species with 38% of total coverage.

5.2.3 Attempt to Quantify

The reads distribution for all markers within either BM or SMs showed a similar exponential pattern as in the pilot study (Figure 8; p.46). Moreover, out of the twenty most abundant species only eight (40%) were present among the most abundant from all datasets, see Table 4; p.53. These findings point out the issues with biomass quantification which is very common among diversity studies.

Table 4. Twenty most abundant genera according to reads numbers and sorted in descending order. Genera present in all datasets are marked with a grey background.

ITS1 Abundance		ITS2 Abundance		*EF-1α_A* Abundance		*RPB2_B* Abundance	
Entoloma	87,48%	*Heterobasidion*	21,94%	*Russula*	25,67%	*Russula*	20,35%
Heterobasidion	4,31%	*Cordyceps*	11,60%	*Cordyceps*	16,31%	*Heterobasidion*	16,71%
Schizophyllum	1,53%	*Melanoleuca*	11,42%	*Lecanicillium*	5,22%	*Melanoleuca*	10,61%
Phoma	0,85%	*Phoma*	9,99%	*Psilocybe*	4,35%	*Lycoperdon*	5,85%
Laetiporus	0,80%	*Lecanicillium*	5,67%	*Laetiporus*	4,02%	*Psilocybe*	4,38%
Russula	0,79%	*Laetiporus*	5,04%	*Umbelopsis*	3,34%	*Cordyceps*	4,01%
Melanoleuca	0,52%	*Psilocybe*	2,46%	*Schizophyllum*	3,01%	*Inonotus*	3,33%
Lecanicillium	0,48%	*Schizophyllum*	2,21%	*Melanoleuca*	2,97%	*Schizophyllum*	3,10%
Pochonia	0,31%	*Entoloma*	2,16%	*Simplicillium*	2,86%	*Hebeloma*	2,81%
Armillaria	0,19%	*Russula*	2,00%	*Galerina*	2,60%	*Phoma*	2,41%
Serpula	0,15%	*Kretzschmaria*	1,70%	*Phoma*	2,48%	*Acremonium*	1,97%
Amanitopsis	0,14%	*Amanitopsis*	1,63%	*Agaricus*	1,55%	*Porodaedalea*	1,95%
Cordyceps	0,14%	*Serpula*	1,40%	*Ischnoderma*	1,40%	*Penicillium*	1,66%
Psilocybe	0,14%	*Penicillium*	1,29%	*Isaria*	1,34%	*Lecanicillium*	1,57%
Torrubiella	0,13%	Pleosporales sp.2	1,17%	*Absidia*	1,31%	*Laetiporus*	1,52%
Isaria	0,11%	*Amanita*	0,93%	*Acremonium*	1,25%	*Agaricus*	1,32%
Petriella	0,10%	*Agaricus*	0,91%	*Trametes*	1,06%	*Mensularia*	1,23%
Mucidula	0,09%	*Trichoderma*	0,84%	*Graphium*	1,06%	*Lepista*	1,18%
Amanita	0,09%	*Agrocybe*	0,80%	*Armillaria*	0,94%	*Rhodocollybia*	0,90%
Cortinarius	0,07%	*Trametes*	0,76%	*Boletus*	0,94%	*Gymnopus*	0,87%

It is difficult to estimate the real cause of these facts. The universality of primers used in environmental studies causes targeting at more potential primer sites. Some of their variants may be amplified preferentially. Noticing this, verified sequences for the proper genes assigned to fungi were trimmed to purely primer sequence and analyzed. In total, there were 268; 106; 1,306 and 530 variants of ITS1; ITS2; *EF-1α_A* and *RPB2_B* for forward primers respectively. Removal of unique haplotypes (singletons) yielded 205; 73; 573 and 295 for ITS1; ITS2; *EF-1α_A* and *RPB2_B* respectively. An ascending order of obtained variants showed moderate dependency of percentage degeneracy within the used

primers on the real GC content in amplified regions; see Figure 15; p.54 as an example for the *RPB2_B* marker. The R^2 value is a coefficient of determination and shows a correlation between the outcomes and their predicted values. Obtained values are not significant, however a dependence of these R^2 values on degeneracy of used primers may be seen; see Figure 16; p.54.

These findings are congruent with other studies that similarly show a preferential amplification of templates with lower GC content (Suzuki & Giovannoni 1996; Pawluczyk *et al.* 2015).

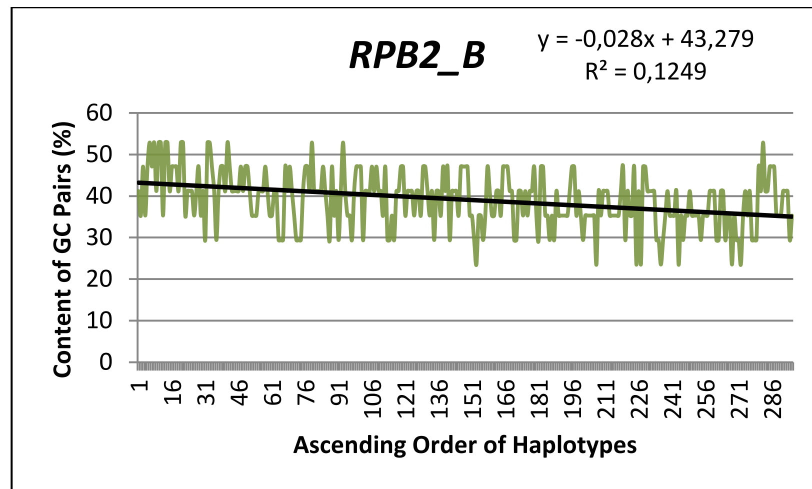

Figure 15. Content of GC pairs within primer sites which were amplified using degenerated primers (29% and 30% degeneracy for fwd. and rev. primers, respectively; shown is the marker *RPB2_B*). The x axis shows an ascending order of haplotypes abundance after singletons removal.

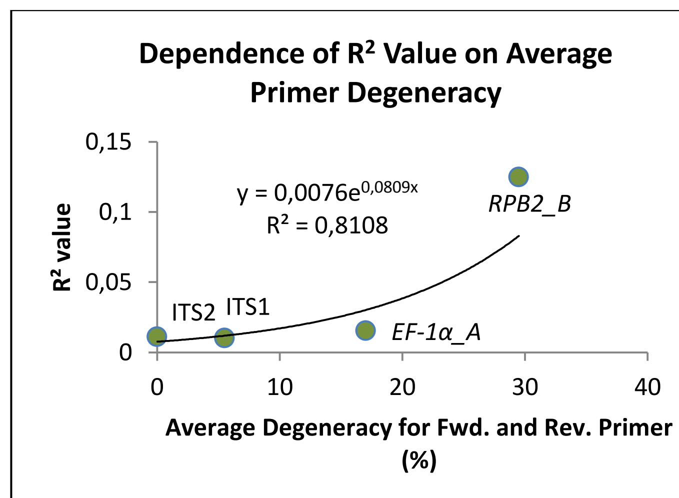

Figure 16. Dependence of R^2 value on average primer degeneracy (calculated from forward and reverse primer values). R^2 value is obtained from the dependency of GC content on the primer site variant abundance. It is clearly seen that higher primer degeneracy (i.e. for *RPB2* marker) leads to a higher R^2 value which can be interpreted as a tendency of highly degenerated primers to preferentially amplify copies with lower GC content in a primer site.

6. Discussion

6.1 Four Primer Combination Exclusion

In total five different markers including eight primer combinations were tested for their potential usability in fungal studies of diversity and ecology. After PCR optimization this selection was reduced to only four markers ITS1, ITS2, *EF-1α* and *RPB2* (when thinking about ITS1 and ITS2 as different markers as they were amplified separately). Four primer combinations were excluded due to inefficient amplification. A plausible explanation for this failure can be found in the fact that incorrect primer combinations were chosen. Despite the enormous volume of literature devoted to selection and design of optimal primers there is no panacea solution and therefore primer combinations that proved inferior were disregarded. There may also have been a problem with input DNA quality. Indeed, an isolation of fungal DNA constitutes a true challenge as there are many PCR inhibiting compounds among fungi (Kosch & Summers 2013). The sample dilution (performed) is a standard procedure leading to an increase of amplification rate. Unfortunately, after these steps DNA could be fragmented or become too diluted to be able to amplify single-copy genes such as *ACT1* or *MCM7*. The amplification of *RPB2* also needed an addition of $MgCl_2$ which significantly improved yields. Only the commonly used primer combination of *EF1α_A* (Rehner & Buckley 2005) led to trouble-free amplification. This may, however, be ascribed to its low-copy occurrence among some taxa (Henk & Fisher 2012). Similarly, the multi-copy marker ITS was amplified successfully in all samples.

6.2 Primers over Marker Selection

The choice of a marker is very important. Nevertheless, in many cases one can't distinguish between influences of the marker itself and its primer combination. Without completely populated databases with sequence data from all markers to be compared one can hardly differentiate between influence of the marker itself and primer combination used. The primer pairs for ITS fragments selected for this study had been chosen according to recent methodological studies (Bellemain *et al.* 2010; Schmidt *et al.* 2013). However, there will still be studies highlighting other combinations, such as Beeck *et al.* (2014). The recovery rate of primers proposed in their study was higher than for other primer pairs recently used. Authors emphasize a broader spectrum of amplified sequences

affiliated to the non-fungal phyla. The question of whether this is the best approach remains, as the broader taxonomic coverage significantly decreases the read proportion for targeted groups. A similar problem can be found for highly degenerated primers which are commonly used for the protein-coding genes. This might be overcome with sufficient sequencing depth (Větrovský *et al.* Submitted). Nonetheless, this is still an unsolved problem as only a few diversity studies reach saturation according to Chao or ACE diversity estimators (reviewed by Zelenka 2013 in Table S2).

6.3 Comparative Analysis of Markers

The main goal of this thesis was to evaluate ITS and alternative markers used in microbial diversity studies. The results presented here show the ITS2 marker as probably the best candidate for environmental studies. It reached the best taxonomic resolution along with best taxonomic coverage. Unfortunately there was an uneven sequencing depth of single gene libraries with the ratio of ITS2 (as the least sequenced marker) to RPB2 reaching 80%. Therefore, resampling for the purpose of marker comparison was set according to ITS2. Nonetheless, even single gene analyses of other markers performed on non-resampled datasets did not alter existing conclusions.

6.3.1 ITS2 versus ITS1

There are many studies arguing which of these markers is better to use. For example Mello *et al.* (2011) assigned more positive features to ITS1. Blaalid *et al.* (2013) put both markers on the same level and Bazzicalupo *et al.* 2013 established ITS2 as more suitable for environmental studies. Our pilot study investigated the ITS1 region and compared it to the *RPB2* gene. *RPB2_P* used in the pilot study proved to have a similar qualitative and better quantitative distribution compared with ITS1. Similarly, in the presented study ITS1 was steadily placed alongside the protein-coding genes, whereas ITS2 proved to be superior from the start of analyzing. The quality of ITS2 is clearly seen from the Venn diagram, (Figure 14; p.52). However, the same figure also visualizes a risk of undiscovered species when only one marker is used. Even though ITS2 revealed 26 extra species; an additional 69 species were observed when other markers were also included in the analysis.

The Chao1 diversity estimator showed different numbers of sequences (ca. 7,000 for ITS2 vs. >20,000 for ITS1) that are needed for recovery of the entire diversity. This is caused by a highly uneven distribution of reads when only a few clusters form a majority,

whereas the rest is represented by small clusters as is the case of ITS1. In diversity estimation this plays a fundamental role as there is a very low probability of uncovering these small groups alongside the major groups. The low probability has to be compensated for by arranging more extensive datasets. This is clearly seen in the example of ITS1 where *Entoloma lividum* occupied 87% of all reads. The exaggerated quantity of *Entoloma* reads were, however, only within the ITS1 dataset. This is difficult to explain as both markers form a tandem array in a genome and their proportions should therefore correspond. One possible explanation could be the different content of GC pairs; therefore this was tested for in both datasets. ITS1 reached 42%±0.0047 and ITS2 showed practically the same percentage. This eliminates the hypothesis that the GC pair content caused the disproportion. Nevertheless, GC content may have an impact on amplification success as was hinted before. Sixteen species were randomly selected from the ITS1 dataset to cover the whole scale of read abundance. There was a slight dependence (R^2=0.1123) of GC content on the read abundance. Thus it may have a certain effect causing *Entoloma* predominance in that particular dataset. Moreover a secondary structure may influence amplification success. The ITS1 sequence is not very predictable and so there are a lack of references. However, the ITS2 structure largely differed among *Entoloma* species within the database of secondary structures (Ankenbrand *et al.* 2015) and differed also when compared with sequences from BM. These structures also differed with respect to values of free energy; in a range of approximately –20 to –75 kcal/mol. DNA sequences of *Entoloma* species that are stored in Genbank varied by up to 15%. Furthermore, Morozova et al. (2014) noticed large insertions in ITS1. All these pieces of knowledge point out a huge heterogeneity within *Entoloma* and unequal amplification can help explain some of the aforementioned issues. Thus, I would recommend the usage of another marker simultaneously with ITS for the study of Entolomataceae.

Overall ITS2 proved to be superior compared with other markers and I recommend it as a barcode marker for environmental studies, in congruence with the recent study of Tedersoo *et al.* (2015).

6.3.2 Protein-Coding Markers

The comparison of protein-coding genes *EF-1α_A* and *RPB2_B* showed a generally similar pattern. *EF-1α_A* reached moderately higher taxonomic coverage than *RPB2_B*. However, *RPB2_B* showed highly correlated relative values of OTUs, read numbers and references, see Table 3; p.49. Unfortunately the low representation of these markers in

public databases could theoretically influence even the creation of reference databases. Nevertheless, the overlapped dataset should serve to compare all markers objectively. In the sense of revealing basal fungal lineages, *RPB2_B* did not confirm results by *RPB2_P* from our pilot study. A total degree of diversity overestimation with 3% level of similarity for clustering analysis was assessed as approximately two. This is slightly bigger than for the *RPB2_P* region. On the other hand the gene *EF-1α_A* showed an extraordinary, almost 8-fold, overestimation. The most interesting results for *EF-1α_A* are attained within the Russulaceae group: the internal transcribed spacer and generally the ribosomal DNA is ubiquitous, spread among all fungi and at the same time this is also a crucial feature of any potential barcode marker. Therefore, usually some house-keeping genes are chosen for this purpose as they are essential for any organism. Apparently this situation is not as clear as it is shown in the example of the Russulaceae family. Results of the query for 'Translation Elongation Factor', '*RPB2'* and 'Internal Transcribed Spacer' in GenBank database demonstrated significant differences. Only two species were assigned to this family for *EF-1α* (less than 1% of hits within the order level), contrary to 1,000 hits for *RPB2* (85% of hits within the order level) and 8,103 hits for ITS (84% of hits within the order level); see Figure 17; p.59.

It is hard to say which taxonomic level can be taken as fundamental. For example *Pholiota* genus from Strophariaceae family is the most abundant genus in the ITS dataset. There are also several hits for the *RPB2* gene, however, no hit for translation elongation factor. And that is despite the fact that *EF-1α* possesses 34% more hits than *RPB2* in total. This is supported by empirical experience from our laboratory where we were not able to amplify any *EF-1α* for the Pholiota genus (unpublished data).

The question remains how is it possible that Russulaceae was included in the reference dataset for *EF-1α_A* given its poor presence in the GenBank database. In SMs there were some sequences recovered and assigned to the Russuales order. In the designed composition of small mocks (containing species from distant phylogenetic groups) one could theoretically unequivocally recognize clusters assigned to the family Russulaceae which suggested that the given sequence was assigned to the *Russula* species. Such inconsistency with the absence of *Russula EF-1α* in the GenBank database might be attributable to noisy Sanger sequences due to the presence of several diverse clusters. Thus, the only way to get sequences from similarly divergent lineages would be either cloning (which is time-consuming and expensive) or high throughput sequencing performed on differentially tagged datasets to verify obtained sequences.

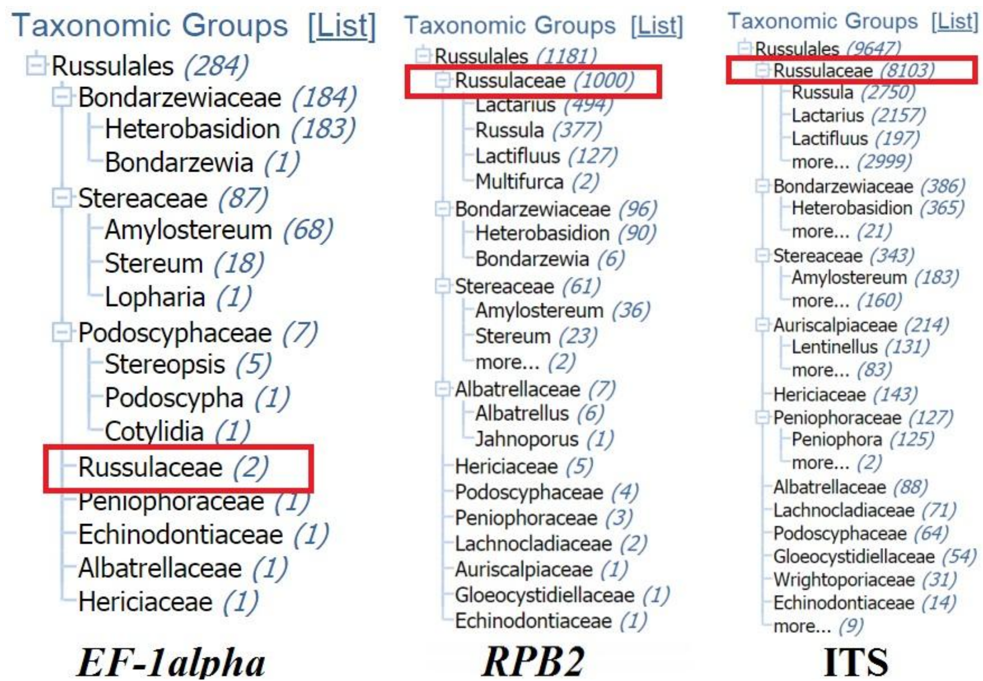

Figure 17. The illustration of poor *EF-1α* representation for Russulaceae family within the GenBank NCBI database over other tested markers *RPB2* and ITS.

Unfortunately, HTS is still not considered to be sufficiently trustworthy to establish novel sequences for species. It is natural when taking into consideration an extreme sensitivity and also all possible contaminants and general uncertainty during the data assignment. Besides Russulaceae there were many other groups showing great variability. Some samples were analyzed for the presence of variability within *EF-1α_A* and several short introns and indels were identified. Moreover, *EF-1α* can exist in more copies within a genome and there are also studies (Keeling & Inagaki 2004; Henk & Fisher 2012; Clouse *et al.* 2013; Kamikawa *et al.* 2013) that mention the existence of *EFL* – 'Elongation Factor Like' within some groups. Furthermore the whole family of closely related proteins around *EF-1α* may serve as a template for degenerate primers and therefore artificially increase the observed diversity. There are fungal groups where both *EF-1α_A* and *RPB2_B* would constitute a powerful tool for analyses. Their usage as self-contained environmental barcode markers is, however, excluded by reasons aforementioned. On the other hand, a high percentage of species might remain undiscovered when only ITS would be used. Accordingly, my general recommendation is to use more markers simultaneously with regard to groups of the presumed community composition.

6.4 Intragenomic Variability

There are studies such as Lindner & Banik (2011) that proved the pronounced overestimation of species number through intragenomic variability. My previous results (Figure 6; p.45 and Figure 7; p.46) confirmed these concerns. However, another study by Lindner *et al.* (2013) essentially denies a significant effect of intragenomic variability within ITS. The minimum reads cutoff for each of their samples was set to 100 reads. This is too low especially when seeing that the average number of ITS copies presumably exceeds 100. With such a low sequencing depth the singleton exclusion (performed) might lead to inaccurate conclusions about the intragenomic variability – even more so when only 28% of clusters were retained in the dataset. Authors even admit a positive correlation between numbers of haplotypes and sequencing depth, nevertheless, they did not directly comment on the possible impact of singleton removal. In this situation one can expect a real risk of a type II error.

The present study confirmed our results from the pilot study when both ITS1 and *RPB2* markers overestimated expected species number. Despite our values from the pilot study: 2.53±1.13 OTUs per species for ITS1 and 1.87±0.92 OTUs for *RPB2_P*; in the present study ITS1 reached slightly lower values than *RPB2_B*. The reason for this inconsistency may be found in the majority of sequences assigned to *Entoloma*, which substantially lowered read proportion of other groups. This could cause insufficient coverage of all intragenomic variants. Another reason may be the use of different primer pairs and therefore different region of *RPB2*. It would thus be necessary to strictly distinguish and more or less change discussions about markers to discussions about primer combinations. Nevertheless, the first hypothesis regarding the potential overestimation of species numbers caused by intragenomic variability within ITS copies seems somewhat exaggerated. Most ITS dissimilarities are probably lower than 3% and thus these sequences cluster together. The number of OTUs is still recorded as being ca. twice overestimated. Taking into consideration the fact of low sequencing depth within some groups, the real value of overestimation may generally be higher than two. The overestimation within the *RPB2_B* dataset was similar to ITS, despite its single-copy character. Likewise as there are known paralogues for other functional genes such as β-tubulin as showed by Hubka & Kolarik (2012), these paralogues might be found for other genes as well. An analysis of sequences from overestimated taxa, performed on *EF-1α_A* and *RPB2_B*, showed a strongly preferred distribution of polymorphism in the third

codon position (data not shown). In the case of sequential errors or pseudogenes these sites would, however, be evenly distributed in sequences. Moreover, the protein alignments reveal that most mutations are silent. Therefore the functional sequences with the same amino acid order can possess a total variability over 3% and thus create more than one cluster.

6.5 Species and Biomass Quantification

One should expect that the comparison of multi-copy and single-copy markers allows a relative estimation of species abundance in environmental samples. In some cases, this can work normally, even though precise methods of quantification such as qPCR must be used. Nevertheless, these attempts might be skewed at several levels. The way of sampling, DNA extraction or just a certain randomness of PCR reaction, all will notably influence the final results. The results presented here showed that in the twenty most abundant species across all markers only 40% of species were present. *Entoloma lividum* in the ITS1 dataset represented about 87% of all reads while the Pearson correlation coefficient of read numbers to OTUs reached 0.9394 and so the general tendency of correlation of these two parameters of diversity estimation is not much affected. Moreover, the pattern of reads distribution shown for the pilot study (Figure 8; p.46) is similar to the pattern seen in the present study (data not shown), both hint at the significant influence of randomness. Taken together with a variable number of ITS copies in a genome (Raidl *et al.* 2005), this creates too many variables that may distort the final results. Thus I don't recommend quantification on the basis of read numbers for either ITS or protein-coding datasets.

7. Conclusions

The present study compared four molecular markers as potential metabarcodes for studies of fungal diversity and ecology. The single gene analysis showed a similar taxonomic coverage with approximately 40% at the species level for ITS1, *EF-1α_A* and *RPB2_B* and almost 55% for ITS2. Next studies were performed on the dataset based on overlapping references. Similarly, the taxonomic resolution analysis evinced ITS2 as the best marker out of those tested. *RPB2_B* showed mostly correlating values of reads, OTUs and observed species and therefore confirmed its effective quantitative representation of community composition observed in the pilot study. Surprisingly only 44 species (16%) were detected by all markers together. This brings attention to usage of a multi-loci approach when *RPB2* could be used alongside the ITS2 as a good alternative marker for studies of microbial diversity.

Ratios of observed OTUs to detected species were compared at species, family and order levels. Generally most overestimated orders belonged to the group Basidiomycota. In total, this analysis revealed enormous (up to 8-fold) overestimation using *EF-1α_A* which practically exclude it from the possibility of being a good metabarcoding marker. Other markers usually oscillated around two, independently of being a multi- or single-copy marker. Yet, there were taxa such as *Entoloma* genus, where the overestimation was substantially higher, e.g. 15-fold for ITS2.

Such huge overestimation can be attributed either to functional or non-functional paralogues within a genome. The ITS region is known for the possibility of intragenomic variability. Nevertheless, such a large distortion for protein-coding genes is surprising. Researching *EF-1α_A* and *RPB2_B* led to the discovery of variability within functional copies. Indeed, most mutations were present in the third codon position which directly rules out the presence of pseudogenes (or sequencing error which have both random distributions). The variability of this site is so high that it can cause a creation of different clusters and therefore possible overestimation. This claim however does not refute the existence of pseudogenes within tested genes but only notifies of the high variability within functional copies.

A general advice for future environmental studies would be to simultaneously employ more distinct markers. When interpreting, one should take into consideration a strong heterogeneity within some groups and also remember that the true diversity revealed in common environmental studies might be overestimated by a factor of 2.

8. References

Anderson, I. C. & Cairney, J. W. G. (2004). Diversity and Ecology of Soil Fungal Communities: Increased Understanding through the Application of Molecular Techniques. Environ. Microbiol. *6*, 769–779.

Ankenbrand, M. J., Keller, A., Wolf, M., Schultz, J. & Förster, F. (2015). ITS2 Database V: Twice as Much. Mol. Biol. Evol. msv174.

Arezi, B., Xing, W., Sorge, J. A. & Hogrefe, H. H. (2003). Amplification Efficiency of Thermostable DNA Polymerases. Anal. Biochem. *321*, 226–235.

Arnold, A. E., Henk, D. A., Eells, R. L., Lutzoni, F. & Vilgalys, R. (2007). Diversity and Phylogenetic Affinities of Foliar Fungal Endophytes in Loblolly Pine Inferred by Culturing and Environmental PCR. Mycologia *99*, 185–206.

Aronesty, E. (2013). Comparison of Sequencing Utility Programs. Open Bioinforma. J. *7*.

Ayliffe, M. A., Dodds, P. N. & Lawrence, G. J. (2001). Characterisation of a *β*-Tubulin Gene from *Melampsora Lini* and Comparison of Fungal *β*-Tubulin Genes. Mycol. Res. *105*, 818–826.

Bailey, C. (2003). Characterization of Angiosperm nrDNA Polymorphism, Paralogy, and Pseudogenes. Mol. Phylogenet. Evol. *29*, 435–455.

Baldrian, P., Kolařík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., Žifčáková, L., Šnajdr, J., Rídl, J., Vlček, Č., *et al.* (2012). Active and Total Microbial Communities in Forest Soil Are Largely Different and Highly Stratified during Decomposition. ISME J. *6*, 248–258.

Baldrian, P., Větrovský, T., Cajthaml, T., Dobiášová, P., Petránková, M., Šnajdr, J. & Eichlerová, I. (2013). Estimation of Fungal Biomass in Forest Litter and Soil. Fungal Ecol. *6*, 1–11.

Barnes, W. M. (1994). PCR Amplification of up to 35-Kb DNA with High Fidelity and High Yield from Lambda Bacteriophage Templates. Proc. Natl. Acad. Sci. U. S. A. *91*, 2216–2220.

Bazzicalupo, A. L., Bálint, M. & Schmitt, I. (2013). Comparison of ITS1 and ITS2 rDNA in 454 Sequencing of Hyperdiverse Fungal Communities. Fungal Ecol. *6*, 102–109.

Beeck, M. Op De, Lievens, B., Busschaert, P., Declerck, S., Vangronsveld, J. & Colpaert, J. V. (2014). Comparison and Validation of Some ITS Primer Pairs Useful for Fungal Metabarcoding Studies. PLoS ONE *9*, e97629.

Begerow, D., Nilsson, H., Unterseher, M. & Maier, W. (2010). Current State and Perspectives of Fungal DNA Barcoding and Rapid Identification Procedures. Appl. Microbiol. Biotechnol. *87*, 99–108.

Belfield, G. P. & Tuite, M. F. (1993). Translation Elongation Factor 3: A Fungus-Specific Translation Factor? Mol. Microbiol. *9*, 411–418.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P. & Kauserud, H. (2010). ITS as an Environmental DNA Barcode for Fungi: An in Silico Approach Reveals Potential PCR Biases. BMC Microbiol. *10*, 189.

Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., de Sousa, F., *et al.* (2013). Improved

Software Detection and Extraction of ITS1 and ITS2 from Ribosomal ITS Sequences of Fungi and Other Eukaryotes for Analysis of Environmental Sequencing Data. Methods Ecol. Evol. *4*, 914–919.

Bjorbækmo, M. F., Carlsen, T., Brysting, A., Vrålstad, T., Høiland, K., Ugland, K. I., Geml, J., Schumacher, T. & Kauserud, H. (2010). High Diversity of Root Associated Fungi in Both Alpine and Arctic Dryas Octopetala. BMC Plant Biol. *10*, 244.

Blaalid, R., Kumar, S., Nilsson, R. H., Abarenkov, K., Kirk, P. M. & Kauserud, H. (2013). ITS1 versus ITS2 as DNA Metabarcodes for Fungi. Mol. Ecol. Resour. *13*, 218–224.

Blackwell, M. (2011). The Fungi: 1, 2, 3 … 5.1 Million Species? Am. J. Bot. *98*, 426–438.

Boon, E., Zimmerman, E., Lang, B. F. & Hijri, M. (2010). Intra-Isolate Genome Variation in Arbuscular Mycorrhizal Fungi Persists in the Transcriptome. J. Evol. Biol. *23*, 1519–1527.

Boyle, B., Hamelin, R. C. & Seguin, A. (2005). In Vivo Monitoring of Obligate Biotrophic Pathogen Growth by Kinetic PCR. Appl. Environ. Microbiol. *71*, 1546–1552.

Brazee, N. J., Hulvey, J. P. & Wick, R. L. (2011). Evaluation of Partial *tef1*, *rpb2*, and nLSU Sequences for Identification of Isolates Representing *Armillaria Calvescens* and *Armillaria Gallica* from Northeastern North America. Fungal Biol. *115*, 741–749.

Brown, S. P., Veach, A. M., Rigdon-Huss, A. R., Grond, K., Lickteig, S. K., Lothamer, K., Oliver, A. K. & Jumpponen, A. (2015). Scraping the Bottom of the Barrel: Are Rare High Throughput Sequences Artifacts? Fungal Ecol. *13*, 221–225.

Buckler, E. S., Ippolito, A. & Holtsford, T. P. (1997). The Evolution of Ribosomal DNA Divergent Paralogues and Phylogenetic Implications. Genetics *145*, 821–832.

Bullerwell, C. E. & Lang, B. F. (2005). Fungal Evolution: The Case of the Vanishing Mitochondrion. Curr. Opin. Microbiol. *8*, 362–369.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., *et al.* (2010). QIIME Allows Analysis of High-Throughput Community Sequencing Data. Nat. Methods *7*, 335–336.

Carlson, A., Justo, A. & Hibbett, D. S. (2014). Species Delimitation in Trametes: A Comparison of ITS, *RPB1*, *RPB2* and *TEF1* Gene Phylogenies. Mycologia *106*, 735–745.

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. Scand. J. Stat. *11*, 265–270.

Clouse, R. M., Sharma, P. P., Giribet, G. & Wheeler, W. C. (2013). Elongation Factor-1 Alpha, a Putative Single-Copy Nuclear Gene, Has Divergent Sets of Paralogs in an Arachnid. Mol. Phylogenet. Evol. *68*, 471–481.

Corradi, N., Croll, D., Colard, A., Kuhn, G., Ehinger, M. & Sanders, I. R. (2007). Gene Copy Number Polymorphisms in an Arbuscular Mycorrhizal Fungal Population. Appl. Environ. Microbiol. *73*, 366–369.

Damon, C., Barroso, G., Férandon, C., Ranger, J., Fraissinet-Tachet, L. & Marmeisse, R. (2010). Performance of the *COX1* Gene as a Marker for the Study of Metabolically Active Pezizomycotina and Agaricomycetes Fungal Communities from the Analysis of Soil RNA. FEMS Microbiol. Ecol. *74*, 693–705.

Daniel, H.-M. & Meyer, W. (2003). Evaluation of Ribosomal RNA and Actin Gene Sequences for the Identification of Ascomycetous Yeasts. Int. J. Food Microbiol. *86*, 61–78.

Daniel, H. M., Sorrell, T. C. & Meyer, W. (2001). Partial Sequence Analysis of the Actin Gene and Its Potential for Studying the Phylogeny of *Candida* Species and Their Teleomorphs. Int. J. Syst. Evol. Microbiol. *51*, 1593–1606.

Debode, J., Van Hemelrijck, W., Baeyen, S., Creemers, P., Heungens, K. & Maes, M. (2009). Quantitative Detection and Monitoring of *Colletotrichum Acutatum* in Strawberry Leaves Using Real-Time PCR. Plant Pathol. *58*, 504–514.

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., Hirsch, P. R. & Vogel, T. M. (2011). Accessing the Soil Metagenome for Studies of Microbial Diversity. Appl. Environ. Microbiol. *77*, 1315–1324.

Dutcher, S. K. (2001). The Tubulin Fraternity: Alpha to Eta. Curr. Opin. Cell Biol. *13*, 49–54.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005). Diversity of the Human Intestinal Microbial Flora. Science *308*, 1635–1638.

Edgar, R. C. (2010). Search and Clustering Orders of Magnitude Faster than BLAST. Bioinforma. Oxf. Engl. *26*, 2460–2461.

Edgar, R. C. (2013). UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads. Nat. Methods *10*, 996–998.

Feau, N., Decourcelle, T., Husson, C., Desprez-Loustau, M.-L. & Dutech, C. (2011). Finding Single Copy Genes out of Sequenced Genomes for Multilocus Phylogenetics in Non-Model Fungi. PLoS ONE *6*, e18803.

Feinstein, L. M., Sul, W. J. & Blackwood, C. B. (2009). Assessment of Bias Associated with Incomplete Extraction of Microbial DNA from Soil. Appl. Environ. Microbiol. *75*, 5428–5433.

Feliner, G. N. & Rosselló, J. A. (2007). Better the Devil You Know? Guidelines for Insightful Utilization of nrDNA ITS in Species-Level Evolutionary Studies in Plants. Mol. Phylogenet. Evol. *44*, 911–919.

Fell, J. W., Boekhout, T., Fonseca, A., Scorzetti, G. & Statzell-Tallman, A. (2000). Biodiversity and Systematics of Basidiomycetous Yeasts as Determined by Large-Subunit rDNA D1/D2 Domain Sequence Analysis. Int. J. Syst. Evol. Microbiol. *50*, 1351–1371.

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R. A., Felts, B., Rayhawk, S., *et al.* (2007). Metagenomic and Small-Subunit rRNA Analyses Reveal the Genetic Diversity of Bacteria, Archaea, Fungi, and Viruses in Soil. Appl. Environ. Microbiol. *73*, 7059–7066.

Frøslev, T. G., Matheny, P. B. & Hibbett, D. S. (2005). Lower Level Relationships in the Mushroom Genus *Cortinarius* (Basidiomycota, Agaricales): A Comparison of *RPB1*, *RPB2*, and ITS Phylogenies. Mol. Phylogenet. Evol. *37*, 602–618.

Ganley, A. R. D. & Kobayashi, T. (2007). Highly Efficient Concerted Evolution in the Ribosomal DNA Repeats: Total rDNA Repeat Variation Revealed by Whole-Genome Shotgun Sequence Data. Genome Res. *17*, 184–191.

Ganley, A. R. D. & Kobayashi, T. (2011). Monitoring the Rate and Dynamics of Concerted Evolution in the Ribosomal DNA Repeats of Saccharomyces Cerevisiae Using Experimental Evolution. Mol. Biol. Evol. *28*, 2883–2891.

Geml, J., Laursen, G. A., Timling, I., Mcfarland, J. M., Booth, M. G., Lennon, N., Nusbaum, C. & Taylor, D. L. (2009). Molecular Phylogenetic Biodiversity Assessment of Arctic and Boreal Ectomycorrhizal *Lactarius* Pers. (Russulales; Basidiomycota) in Alaska, Based on Soil and Sporocarp DNA. Mol. Ecol. *18*, 2213–2227.

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. Mol. Phylogenet. Evol. *67*, 234–245.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., *et al.* (2011). Chimeric 16S rRNA Sequence Formation and Detection in Sanger and 454-Pyrosequenced PCR Amplicons. Genome Res. *21*, 494–504.

Hansen, K., LoBuglio, K. F. & Pfister, D. H. (2005). Evolutionary Relationships of the Cup-Fungus Genus *Peziza* and Pezizaceae Inferred from Multiple Nuclear Genes: *RPB2*, *β*-Tubulin, and LSU rDNA. Mol. Phylogenet. Evol. *36*, 1–23.

Hao, X., Jiang, R. & Chen, T. (2011). Clustering 16S rRNA for OTU Prediction: A Method of Unsupervised Bayesian Clustering. Bioinformatics.

Hartmann, M., Howes, C. G., VanInsberghe, D., Yu, H., Bachar, D., Christen, R., Henrik Nilsson, R., Hallam, S. J. & Mohn, W. W. (2012). Significant and Persistent Impact of Timber Harvesting on Soil Microbial Communities in Northern Coniferous Forests. ISME J. *6*, 2320–2320.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. (2003). Biological Identifications through DNA Barcodes. Proc. R. Soc. B Biol. Sci. *270*, 313–321.

Henk, D. A. & Fisher, M. C. (2012). The Gut Fungus *Basidiobolus Ranarum* Has a Large Genome and Different Copy Numbers of Putatively Functionally Redundant Elongation Factor Genes. PLoS ONE *7*, e31268.

Henk, D. A., Weir, A. & Blackwell, M. (2003). *Laboulbeniopsis Termitarius*, an Ectoparasite of Termites Newly Recognized as a Member of the Laboulbeniomycetes. Mycologia *95*, 561–564.

Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., Huhndorf, S., James, T., Kirk, P. M., Lücking, R., *et al.* (2007). A Higher-Level Phylogenetic Classification of the Fungi. Mycol. Res. *111*, 509–547.

Hibbett, D. S., Ohman, A., Glotzer, D., Nuhn, M., Kirk, P. & Nilsson, R. H. (2011). Progress in Molecular and Morphological Taxon Discovery in Fungi and Options for Formal Classification of Environmental Sequences. Fungal Biol. Rev. *25*, 38–47.

Hirt, R. P., Logsdon, J. M., Healy, B., Dorey, M. W., Doolittle, W. F. & Embley, T. M. (1999). Microsporidia Are Related to Fungi: Evidence from the Largest Subunit of RNA Polymerase II and Other Proteins. Proc. Natl. Acad. Sci. U. S. A. *96*, 580–585.

Hoffmann, K., Pawłowska, J., Walther, G., Wrzosek, M., de Hoog, G. S., Benny, G. L., Kirk, P. M. & Voigt, K. (2013). The Family Structure of the Mucorales: A Synoptic Revision Based on Comprehensive Multigene-Genealogies. Persoonia - Mol. Phylogeny Evol. Fungi *30*, 57–76.

Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S., Erickson, D. L., Fazekas, A. J., *et al.* (2009). A DNA Barcode for Land Plants. Proc. Natl. Acad. Sci. *106*, 12794–12797.

Huang, C.-H., Chang, M.-T. & Huang, L. (2012). Species Identification of *Wickerhamomyces Anomalus* and Related Taxa Using *β*-Tubulin (*β-Tub*) DNA Barcode Marker. Yeast *29*, 531–535.

Hubka, V. & Kolarik, M. (2012). *β*-Tubulin Paralogue *tubC* Is Frequently Misidentified as the *benA* Gene in *Aspergillus* Section *Nigri* Taxonomy: Primer Specificity Testing and Taxonomic Consequences. Persoonia Mol. Phylogeny Evol. Fungi *29*, 1–10.

Hwang, J., Zhao, Q., Yang, Z. L., Wang, Z. & Townsend, J. P. (2015). Solving the Ecological Puzzle of Mycorrhizal Associations Using Data from Annotated Collections and Environmental Samples – an Example of Saddle Fungi. Environ. Microbiol. Rep. *7*, 658–667.

Ide, S., Miyazaki, T., Maki, H. & Kobayashi, T. (2010). Abundance of Ribosomal RNA Gene Copies Maintains Genome Integrity. Science *327*, 693–696.

Ihrmark, K., Bödeker, I. T. M., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., Clemmensen, K. E., *et al.* (2012). New Primers to Amplify the Fungal ITS2 Region – Evaluation by 454-Sequencing of Artificial and Natural Communities. FEMS Microbiol. Ecol. *82*, 666–677.

James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., *et al.* (2006). Reconstructing the Early Evolution of Fungi Using a Six-Gene Phylogeny. Nature *443*, 818–822.

Jumpponen, A. (2003). Soil Fungal Community Assembly in a Primary Successional Glacier Forefront Ecosystem as Inferred from rDNA Sequence Analyses. New Phytol. *158*, 569–578.

Jumpponen, A., Jones, K. L. & Blair, J. (2010). Vertical Distribution of Fungal Communities in Tallgrass Prairie Soil. Mycologia *102*, 1027–1041.

Kamikawa, R., Brown, M. W., Nishimura, Y., Sako, Y., Heiss, A. A., Yubuki, N., Gawryluk, R., Simpson, A. G., Roger, A. J., Hashimoto, T., *et al.* (2013). Parallel Re-Modeling of *EF-1α* Function: Divergent *EF-1α* Genes Co-Occur with *EFL* Genes in Diverse Distantly Related Eukaryotes. BMC Evol. Biol. *13*, 131.

Kauserud, H., Colman, J. E. & Ryvarden, L. (2008). Relationship between Basidiospore Size, Shape and Life History Characteristics: A Comparison of Polypores. Fungal Ecol. *1*, 19–23.

Keeling, P. J. & Inagaki, Y. (2004). A Class of Eukaryotic GTPase with a Punctate Distribution Suggesting Multiple Functional Replacements of Translation Elongation Factor 1*α*. Proc. Natl. Acad. Sci. U. S. A. *101*, 15380–15385.

Kiss, L. (2012). Limits of Nuclear Ribosomal DNA Internal Transcribed Spacer (ITS) Sequences as Species Barcodes for Fungi. Proc. Natl. Acad. Sci. *109*, E1811–E1811.

Ko, K. S. & Jung, H. S. (2002). Three Nonorthologous ITS1 Types Are Present in a Polypore Fungus *Trichaptum Abietinum*. Mol. Phylogenet. Evol. *23*, 112–122.

Kobayashi, T. (2011). Regulation of Ribosomal RNA Gene Copy Number and Its Role in Modulating Genome Integrity and Evolutionary Adaptability in Yeast. Cell. Mol. Life Sci. *68*, 1395–1403.

Kohout, P., Sudová, R., Janoušková, M., Čtvrtlíková, M., Hejda, M., Pánková, H., Slavíková, R., Štajerová, K., Vosátka, M. & Sýkorová, Z. (2014). Comparison of Commonly Used Primer Sets for Evaluating Arbuscular Mycorrhizal Fungal Communities: Is There a Universal Solution? Soil Biol. Biochem. *68*, 482–493.

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., *et al.* (2013). Towards a Unified Paradigm for Sequence-Based Identification of Fungi. Mol. Ecol. *22*, 5271–5277.

Kosch, T. A. & Summers, K. (2013). Techniques for Minimizing the Effects of PCR Inhibitors in the Chytridiomycosis Assay. Mol. Ecol. Resour. *13*, 230–236.

Kouvelis, V. N., Ghikas, D. V. & Typas, M. A. (2004). The Analysis of the Complete Mitochondrial Genome of *Lecanicillium Muscarium* (synonym *Verticillium Lecanii*) Suggests a Minimum Common Gene Organization in mtDNAs of Sordariomycetes: Phylogenetic Implications. Fungal Genet. Biol. *41*, 930–940.

Kovacs, G. M., Jankovics, T. & Kiss, L. (2011). Variation in the nrDNA ITS Sequences of Some Powdery Mildew Species: Do Routine Molecular Identification Procedures Hide Valuable Information? Eur. J. Plant Pathol. *131*, 135–141.

Krak, K., Janoušková, M., Caklová, P., Vosátka, M. & Štorchová, H. (2012). Intraradical Dynamics of Two Coexisting Isolates of the Arbuscular Mycorrhizal Fungus *Glomus* Intraradices *sensu Lato* as Estimated by Real-Time PCR of Mitochondrial DNA. Appl. Environ. Microbiol. *78*, 3630–3637.

Kretzer, A. M. & Bruns, T. D. (1999). Use of *atp6* in Fungal Phylogenetics: An Example from the Boletales. Mol. Phylogenet. Evol. *13*, 483–492.

Kunin, V., Engelbrektson, A., Ochman, H. & Hugenholtz, P. (2010). Wrinkles in the Rare Biosphere: Pyrosequencing Errors Can Lead to Artificial Inflation of Diversity Estimates. Environ. Microbiol. *12*, 118–123.

Lafontaine, I. & Dujon, B. (2010). Origin and Fate of Pseudogenes in Hemiascomycetes: A Comparative Analysis. BMC Genomics *11*, 260.

Lehmann, J. & Libchaber, A. (2008). Degeneracy of the Genetic Code and Stability of the Base Pair at the Second Position of the Anticodon. RNA *14*, 1264–1269.

Lim, Y. W., Alamouti, S. M., Kim, J.-J., Lee, S. & Breuil, C. (2004). Multigene Phylogenies of *Ophiostoma Clavigerum* and Closely Related Species from Bark Beetle-Attacked Pinus in North America. FEMS Microbiol. Lett. *237*, 89–96.

Lindner, D. L. & Banik, M. T. (2011). Intragenomic Variation in the ITS rDNA Region Obscures Phylogenetic Relationships and Inflates Estimates of Operational Taxonomic Units in Genus Laetiporus. Mycologia *103*, 731–740.

Lindner, D. L., Carlsen, T., Henrik Nilsson, R., Davey, M., Schumacher, T. & Kauserud, H. (2013). TUDLE.Employing 454 Amplicon Pyrosequencing to Reveal Intragenomic

Divergence in the Internal Transcribed Spacer rDNA Region in Fungi. Ecol. Evol. *3*, 1751–1764.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L. & Law, M. (2012). Comparison of Next-Generation Sequencing Systems. J. Biomed. Biotechnol. *2012*.

Liu, Y. J., Whelen, S. & Hall, B. D. (1999). Phylogenetic Relationships among Ascomycetes: Evidence from an RNA Polymerse II Subunit. Mol. Biol. Evol. *16*, 1799–1808.

Lopez, J. V., Yuhki, N., Masuda, R., Modi, W. & O'Brien, S. J. (1994). Numt, a Recent Transfer and Tandem Amplification of Mitochondrial DNA to the Nuclear Genome of the Domestic Cat. J. Mol. Evol. *39*, 174–190.

Lutzoni, F., Kauff, F., Cox, C. J., McLaughlin, D., Celio, G., Dentinger, B., Padamsee, M., Hibbett, D., James, T. Y., Baloch, E., *et al.* (2004). Assembling the Fungal Tree of Life: Progress, Classification, and Evolution of Subcellular Traits. Am. J. Bot. *91*, 1446–1480.

Mahé, S., Duhamel, M., Calvez, T. L., Guillot, L., Sarbu, L., Bretaudeau, A., Collin, O., Dufresne, A., Kiers, E. T. & Vandenkoornhuyse, P. (2012). PHYMYCO-DB: A Curated Database for Analyses of Fungal Diversity and Evolution. PLOS ONE *7*, e43117.

Martin, K. J. & Rygiewicz, P. T. (2005). Fungal-Specific PCR Primers Developed for Analysis of the ITS Region of Environmental DNA Extracts. BMC Microbiol. *5*, 28.

Matheny, P. B. (2005). Improving Phylogenetic Inference of Mushrooms with *RPB1* and *RPB2* Nucleotide Sequences (*Inocybe*; Agaricales). Mol. Phylogenet. Evol. *35*, 1–20.

Matheny, P. B., Curtis, J. M., Hofstetter, V., Aime, M. C., Moncalvo, J.-M., Ge, Z.-W., Yang, Z.-L., Slot, J. C., Ammirati, J. F., Baroni, T. J., *et al.* (2006). Major Clades of Agaricales: A Multilocus Phylogenetic Overview. Mycologia *98*, 982–995.

Mayol, M. & Rosselló, J. A. (2001). Why Nuclear Ribosomal DNA Spacers (ITS) Tell Different Stories in Quercus. Mol. Phylogenet. Evol. *19*, 167–176.

Mello, A., Napoli, C., Murat, C., Morin, E., Marceddu, G. & Bonfante, P. (2011). ITS-1 versus ITS-2 Pyrosequencing: A Comparison of Fungal Populations in Truffle Grounds. Mycologia *103*, 1184–1193.

Mirhendi, H., Makimura, K., Hoog, G. S. de, Rezaei-Matehkolaei, A., Najafzadeh, M. J., Umeda, Y. & Ahmadi, B. (2014). Translation Elongation Factor 1-α Gene as a Potential Taxonomic and Identification Marker in Dermatophytes. Med. Mycol. myu088.

Molitor, C., Inthavong, B., Sage, L., Geremia, R. A. & Mouhamadou, B. (2010). Potentiality of the *cox1* Gene in the Taxonomic Resolution of Soil Fungi. FEMS Microbiol. Lett. *302*, 76–84.

Monard, C., Gantner, S. & Stenlid, J. (2013). Utilizing ITS1 and ITS2 to Study Environmental Fungal Diversity Using Pyrosequencing. FEMS Microbiol. Ecol. *84*, 165–175.

Morgenstern, I., Powlowski, J., Ishmael, N., Darmond, C., Marqueteau, S., Moisan, M.-C., Quenneville, G. & Tsang, A. (2012). A Molecular Phylogeny of Thermophilic Fungi. Fungal Biol. *116*, 489–502.

Morozova, O. V., Noordeloos, M. E. & Vila, J. (2014). Entoloma Subgenus Leptonia in Boreal-Temperate Eurasia: Towards a Phylogenetic Species Concept. Persoonia Mol. Phylogeny Evol. Fungi *32*, 141–169.

Mouhamadou, B., Carriconde, F., Gryta, H., Jargeat, P., Manzi, S. & Gardes, M. (2008). Molecular Evolution of Mitochondrial Ribosomal DNA in the Fungal Genus *Tricholoma*: Barcoding Implications. Fungal Genet. Biol. *45*, 1219–1226.

Nilsson, R. H., Kristiansson, E., Ryberg, M., Hallenberg, N. & Larsson, K.-H. (2008). Intraspecific ITS Variability in the Kingdom Fungi as Expressed in the International Sequence Databases and Its Implications for Molecular Species Identification. Evol. Bioinforma. Online *4*, 193–201.

O'Brien, H. E., Parrent, J. L., Jackson, J. A., Moncalvo, J.-M. & Vilgalys, R. (2005). Fungal Community Analysis by Large-Scale Sequencing of Environmental Samples. Appl. Environ. Microbiol. *71*, 5544–5550.

O'Donnell, K. & Cigelnik, E. (1997). Two Divergent Intragenomic rDNA ITS2 Types within a Monophyletic Lineage of the Fungus Fusarium Are Nonorthologous. Mol. Phylogenet. Evol. *7*, 103–116.

Oliver, A. K., Brown, S. P., Callaham Jr., M. A. & Jumpponen, A. (2015). Polymerase Matters: Non-Proofreading Enzymes Inflate Fungal Community Richness Estimates by up to 15 %. Fungal Ecol. *15*, 86–89.

Öpik, M., Metsis, M., Daniell, T. J., Zobel, M. & Moora, M. (2009). Large-Scale Parallel 454 Sequencing Reveals Host Ecological Group Specificity of Arbuscular Mycorrhizal Fungi in a Boreonemoral Forest. New Phytol. *184*, 424–437.

Orgiazzi, A., Bianciotto, V., Bonfante, P., Daghino, S., Ghignone, S., Lazzari, A., Lumini, E., Mello, A., Napoli, C., Perotto, S., *et al.* (2013). 454 Pyrosequencing Analysis of Fungal Assemblages from Geographically Distant, Disparate Soils Reveals Spatial Patterning and a Core Mycobiome. Diversity *5*, 73–98.

Ota, Y., Tokuda, S., Buchanan, P. K. & Hattori, T. (2006). Phylogenetic Relationships of Japanese Species of *Heterobasidion—H. Annosum* Sensu Lato and an Undetermined *Heterobasidion Sp.* Mycologia *98*, 717–725.

Pantou, M. P., Kouvelis, V. N. & Typas, M. A. (2006). The Complete Mitochondrial Genome of the Vascular Wilt Fungus *Verticillium Dahliae*: A Novel Gene Order for <I>Verticillium</i> and a Diagnostic Tool for Species Identification. Curr. Genet. *50*, 125–136.

Pawluczyk, M., Weiss, J., Links, M. G., Aranguren, M. E., Wilkinson, M. D. & Egea-Cortines, M. (2015). Quantitative Evaluation of Bias in PCR Amplification and Next-Generation Sequencing Derived from Metabarcoding Samples. Anal. Bioanal. Chem. *407*, 1841–1848.

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P. & Gu, Y. (2012). A Tale of Three next Generation Sequencing Platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq Sequencers. BMC Genomics *13*, 341.

Raab, P. A., Brennwald, A. & Redecker, D. (2005). Mitochondrial Large Ribosomal Subunit Sequences Are Homogeneous within Isolates of *Glomus* (arbuscular Mycorrhizal Fungi, Glomeromycota). Mycol. Res. *109*, 1315–1322.

Raidl, S., Bonfigli, R. & Agerer, R. (2005). Calibration of Quantitative Real-Time Taqman PCR by Correlation with Hyphal Biomass and ITS Copies in Mycelia of *Piloderma Croceum*. Plant Biol. *7*, 713–717.

Raja, H., Schoch, C. L., Hustad, V., Shearer, C. & Miller, A. (2011). Testing the Phylogenetic Utility of *MCM7* in the Ascomycota. MycoKeys *1*, 63–94.

Rajashekar, B., Samson, P., Johansson, T. & Tunlid, A. (2007). Evolution of Nucleotide Sequences and Expression Patterns of Hydrophobin Genes in the Ectomycorrhizal Fungus Paxillus Involutus. New Phytol. *174*, 399–411.

Rappé, M. S. & Giovannoni, S. J. (2003). The Uncultured Microbial Majority. Annu. Rev. Microbiol. *57*, 369–394.

Redecker, D. & Raab, P. (2006). Phylogeny of the Glomeromycota (arbuscular Mycorrhizal Fungi): Recent Developments and New Gene Markers. Mycologia *98*, 885–895.

Reeb, V., Lutzoni, F. & Roux, C. (2004). Contribution of RPB2 to Multilocus Phylogenetic Studies of the Euascomycetes (Pezizomycotina, Fungi) with Special Emphasis on the Lichen-Forming Acarosporaceae and Evolution of Polyspory. Mol. Phylogenet. Evol. *32*, 1036–1060.

Rehner, S. A. & Buckley, E. (2005). A *Beauveria* Phylogeny Inferred from Nuclear ITS and *EF1-α* Sequences: Evidence for Cryptic Diversification and Links to *Cordyceps* Teleomorphs. Mycologia *97*, 84–98.

Roger, A. J., Sandblom, O., Doolittle, W. F. & Philippe, H. (1999). An Evaluation of Elongation Factor 1 Alpha as a Phylogenetic Marker for Eukaryotes. Mol. Biol. Evol. *16*, 218–233.

Rooney, A. P. & Ward, T. J. (2005). Evolution of a Large Ribosomal RNA Multigene Family in Filamentous Fungi: Birth and Death of a Concerted Evolution Paradigm. Proc. Natl. Acad. Sci. U. S. A. *102*, 5084–5089.

Rydholm, C., Dyer, P. S. & Lutzoni, F. (2007). DNA Sequence Characterization and Molecular Evolution of MAT1 and MAT2 Mating-Type Loci of the Self-Compatible Ascomycete Mold Neosartorya Fischeri. Eukaryot. Cell *6*, 868–874.

Sadowska-Deś, A. D., Bálint, M., Otte, J. & Schmitt, I. (2013). Assessing Intraspecific Diversity in a Lichen-Forming Fungus and Its Green Algal Symbiont: Evaluation of Eight Molecular Markers. Fungal Ecol. *6*, 141–151.

Samson, R. A., Seifert, K. A., Kuijpers, A. F., Houbraken, J. & Frisvad, J. C. (2004). Phylogenetic Analysis of *Penicillium* Subgenus *Penicillium* Using Partial *β*-Tubulin Sequences. Stud Mycol *49*, 175–200.

Schander, C. & Willassen, E. (2005). What Can Biological Barcoding Do for Marine Biology? Mar. Biol. Res. *1*, 79–83.

Schmidt, P.-A., Balint, M., Greshake, B., Bandow, C., Roembke, J. & Schmitt, I. (2013). Illumina Metabarcoding of a Soil Fungal Community. Soil Biol. Biochem. *65*, 128–132.

Schmitt, I., Crespo, A., Divakar, P. K., Fankhauser, J. D., Herman-Sackett, E., Kalb, K., Nelsen, M. P., Nelson, N. A., Rivas-Plata, E., Shimp, A. D., *et al.* (2009). New Primers for Promising Single-Copy Genes in Fungal Phylogenetics and Systematics. Persoonia - Mol. Phylogeny Evol. Fungi *23*, 35–40.

Schoch, C. L., Robbertse, B., Robert, V., Vu, D., Cardinali, G., Irinyi, L., Meyer, W., Nilsson, R. H., Hughes, K., Miller, A. N., *et al.* (2014). Finding Needles in Haystacks: Linking Scientific Names, Reference Specimens and Molecular Data for Fungi. Database J. Biol. Databases Curation *2014*.

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., *et al.* (2012). Nuclear Ribosomal Internal Transcribed Spacer (ITS) Region as a Universal DNA Barcode Marker for Fungi. Proc. Natl. Acad. Sci. U. S. A. *109*, 6241–6246.

Scorzetti, G., Fell, J. W., Fonseca, A. & Statzell-Tallman, A. (2002). Systematics of Basidiomycetous Yeasts: A Comparison of Large Subunit D1/D2 and Internal Transcribed Spacer rDNA Regions. FEMS Yeast Res. *2*, 495–517.

Seifert, K. A. (2009). Progress towards DNA Barcoding of Fungi. Mol. Ecol. Resour. *9*, 83–89.

Seifert, K. A., Samson, R. A., deWaard, J. R., Houbraken, J., Lévesque, C. A., Moncalvo, J.-M., Louis-Seize, G. & Hebert, P. D. N. (2007). Prospects for Fungus Identification Using *CO1* DNA Barcodes, with *Penicillium* as a Test Case. Proc. Natl. Acad. Sci. *104*, 3901–3906.

Shi, T.-L. & Perlin, M. H. (2001). The *β*-Tubulin-Encoding Gene from *Microbotryum Violaceum*: Unusual in a Variety of Ways. Curr. Genet. *39*, 253–263.

Shimono, Y., Kato, M. & Takamatsu, S. (2004). Molecular Phylogeny of Russulaceae (Basidiomycetes; Russulales) Inferred from the Nucleotide Sequences of Nuclear Large Subunit rDNA. Mycoscience *45*, 303–316.

da Silva, G. A., Lumini, E., Maia, L. C., Bonfante, P. & Bianciotto, V. (2006). Phylogenetic Analysis of Glomeromycota by Partial LSU rDNA Sequences. Mycorrhiza *16*, 183–189.

Simon, U. K. & Weiss, M. (2008). Intragenomic Variation of Fungal Ribosomal Genes Is Higher than Previously Thought. Mol. Biol. Evol. *25*, 2251–2254.

Smith, M. E., Douhan, G. W. & Rizzo, D. M. (2007). Intra-Specific and Intra-Sporocarp ITS Variation of Ectomycorrhizal Fungi as Assessed by rDNA Sequencing of Sporocarps and Pooled Ectomycorrhizal Roots from a *Quercus* Woodland. Mycorrhiza *18*, 15–22.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M. & Herndl, G. J. (2006). Microbial Diversity in the Deep Sea and the Underexplored "rare Biosphere." Proc. Natl. Acad. Sci. *103*, 12115–12120.

Sosa-Gómez, D. R., Humber, R. A., Hodge, K. T., Binneck, E. & da Silva-Brandão, K. L. (2008). Variability of the Mitochondrial SSU rDNA of Nomuraea Species and Other Entomopathogenic Fungi from Hypocreales. Mycopathologia *167*, 145–154.

Stockinger, H., Peyret-Guzzon, M., Koegel, S., Bouffaud, M.-L. & Redecker, D. (2014). The Largest Subunit of RNA Polymerase II as a New Marker Gene to Study Assemblages of Arbuscular Mycorrhizal Fungi in the Field. PLoS ONE *9*, e107783.

Suzuki, M. T. & Giovannoni, S. J. (1996). Bias Caused by Template Annealing in the Amplification of Mixtures of 16S rRNA Genes by PCR. Appl. Environ. Microbiol. *62*, 625–630.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. & Willerslev, E. (2012). Towards Next-Generation Biodiversity Assessment Using DNA Metabarcoding. Mol. Ecol. *21*, 2045–2050.

Tanabe, Y., Saikawa, M., Watanabe, M. M. & Sugiyama, J. (2004). Molecular Phylogeny of Zygomycota Based on *EF-1α* and *RPB1* Sequences: Limitations and Utility of Alternative Markers to rDNA. Mol. Phylogenet. Evol. *30*, 438–449.

Tang, A. M. C., Jeewon, R. & Hyde, K. D. (2006). Phylogenetic Utility of Protein (*RPB2*, *β*-Tubulin) and Ribosomal (LSU, SSU) Gene Sequences in the Systematics of Sordariomycetes (Ascomycota, Fungi). Antonie Van Leeuwenhoek *91*, 327–349.

Tarkka, M. T., Vasara, R., Gorfer, M. & Raudaskoski, M. (2000). Molecular Characterization of Actin Genes from Homobasidiomycetes: Two Different Actin Genes from *Schizophyllum Commune* and *Suillus Bovinus*. Gene *251*, 27–35.

Taylor, D. L., Booth, M. G., Mcfarland, J. W., Herriott, I. C., Lennon, N. J., Nusbaum, C. & Marr, T. G. (2008). Increasing Ecological Inference from High Throughput Sequencing of Fungi in the Environment through a Tagging Approach. Mol. Ecol. Resour. *8*, 742–752.

Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., Kõljalg, U., Kisand, V., Nilsson, H., Hildebrand, F., *et al.* (2015). Shotgun Metagenomes and Multiple Primer Pair-Barcode Combinations of Amplicons Reveal Biases in Metabarcoding Analyses of Fungi. MycoKeys *10*, 1–43.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V., Vasco-Palacios, A. M., Thu, P. Q., Suija, A., *et al.* (2014). Global Diversity and Geography of Soil Fungi. Science *346*, 1256688.

Tedersoo, L., Nilsson, R. H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., Bahram, M., Bechem, E., Chuyong, G. & Kõljalg, U. (2010). 454 Pyrosequencing and Sanger Sequencing of Tropical Mycorrhizal Fungi Provide Similar Results but Reveal Substantial Methodological Biases. New Phytol. *188*, 291–301.

Thakuria, D., Schmidt, O., Mac Siúrtáin, M., Egan, D. & Doohan, F. M. (2008). Importance of DNA Quality in Comparative Soil Microbial Community Structure Analyses. Soil Biol. Biochem. *40*, 1390–1403.

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*, e40863.

Tomšovský, M., Menkis, A. & Vasaitis, R. (2010). Phylogenetic Relationships in European *Ceriporiopsis* Species Inferred from Nuclear and Mitochondrial Ribosomal DNA Sequences. Fungal Biol. *114*, 350–358.

Tretter, E. D., Johnson, E. M., Wang, Y., Kandel, P. & White, M. M. (2013). Examining New Phylogenetic Markers to Uncover the Evolutionary History of Early-Diverging Fungi: Comparing *MCM7*, *TSR1* and rRNA Genes for Single- and Multi-Gene Analyses of the Kickxellomycotina. Persoonia Mol. Phylogeny Evol. Fungi *30*, 106–125.

Truong, C., Divakar, P. K., Yahr, R., Crespo, A. & Clerc, P. (2013). Testing the Use of ITS rDNA and Protein-Coding Genes in the Generic and Species Delimitation of the Lichen Genus *Usnea*(Parmeliaceae, Ascomycota). Mol. Phylogenet. Evol. *68*, 357–372.

Unterseher, M., Jumpponen, A., Öpik, M., Tedersoo, L., Moora, M., Dormann, C. F. & Schnittler, M. (2011). Species Abundance Distributions and Richness Estimations in Fungal Metagenomics – Lessons Learned from Community Ecology. Mol. Ecol. *20*, 275–285.

Vela-Corcía, D., Bellón-Gómez, D., López-Ruiz, F., Torés, J. A. & Pérez-García, A. (2014). The *Podosphaera Fusca TUB2* Gene, a Molecular “Swiss Army Knife” with Multiple Applications in Powdery Mildew Research. Fungal Biol. *118*, 228–241.

Veselská, T. & Kolařík, M. (2015). Application of Flow Cytometry for Exploring the Evolution of *Geosmithia* Fungi Living in Association with Bark Beetles: The Role of Conidial DNA Content. Fungal Ecol. *13*, 83–92.

Vestheim, H. & Jarman, S. N. (2008). Blocking Primers to Enhance PCR Amplification of Rare Sequences in Mixed Samples – a Case Study on Prey DNA in Antarctic Krill Stomachs. Front. Zool. *5*, 12.

Větrovský, T. & Baldrian, P. (2013). Analysis of Soil Fungal Communities by Amplicon Pyrosequencing: Current Approaches to Data Analysis and the Introduction of the Pipeline SEED. Biol. Fertil. Soils 1–11.

Větrovský, T., Kolařík, M., Žifčáková, L., Zelenka, T. & Baldrian, P. (Submitted). The *rpb2* Gene Represents a Viable Alternative Molecular Marker for the Analysis of Environmental Fungal Communities. Mol. Ecol. Resour.

Vialle, A., Feau, N., Allaire, M., Didukh, M., Martin, F., Moncalvo, J.-M. & Hamelin, R. C. (2009). Evaluation of Mitochondrial Genes as DNA Barcode for Basidiomycota. Mol. Ecol. Resour. *9*, 99–113.

Voglmayr, H., Choi, Y.-J. & Shin, H.-D. (2013). Multigene Phylogeny, Taxonomy and Reclassification of *Hyaloperonospora* on Cardamine. Mycol. Prog. *13*, 131–144.

Voigt, K. & Wöstemeyer, J. (2000). Reliable Amplification of Actin Genes Facilitates Deep-Level Phylogeny. Microbiol. Res. *155*, 179–195.

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. Appl. Environ. Microbiol. *73*, 5261–5267.

White, T., Bruns, T., Lee, S. & Taylor, J. (1990). Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. PCR Protoc. Guide Methods Appl. 315–322.

Wiemers, M. & Fiedler, K. (2007). Does the DNA Barcoding Gap Exist? – a Case Study in Blue Butterflies (*Lepidoptera*: Lycaenidae). Front. Zool. *4*, 8.

Will, K. W. & Rubinoff, D. (2004). Myth of the Molecule: DNA Barcodes for Species Cannot Replace Morphology for Identification and Classification. Cladistics *20*, 47–55.

Wubet, T., Christ, S., Schöning, I., Boch, S., Gawlich, M., Schnabel, B., Fischer, M. & Buscot, F. (2012). Differences in Soil Fungal Communities between European Beech (*Fagus Sylvatica* L.) Dominated Forests Are Related to Soil and Understory Vegetation. PloS One *7*, e47500.

Zelenka, T. (2013). Studium Environmentálních Vzorků Pomocí Pyrosekvenování a Vliv Intragenomové Variability ITS rDNA Oblasti Na Odhady Diverzity Hub. Bakalářská Práce Přír. Fak. Univerzity Karlovy V Praze Kated. Bot. Ved. Bakalářské Práce Miroslav Kolařík.

Zeng, Q.-Y., Rasmuson-Lestander, A. & Wang, X.-R. (2004). Extensive Set of Mitochondrial LSU rDNA-Based Oligonucleotide Probes for the Detection of Common Airborne Fungi. FEMS Microbiol. Lett. *237*, 79–87.

Zhan, A. & MacIsaac, H. J. (2014). Rare Biosphere Exploration Using High-Throughput Sequencing: Research Progress and Perspectives. Conserv. Genet. *16*, 513–522.

9. Supplementary

9.1 Supplementary Table 1

Table S1. The summary of all items included in 'Big' Mock community. More detailed information about reference sequences and SM composition is provided in Table S2 (Attached on CD)
Legend:
No. = Number of identical species included in dataset
Phyl. (Subphyl) = Phylum or Subphylum in basal fungal lineages; Asc., Bas., Kic., Mor. and Muc. denote groups of Ascomycota, Basidiomycota, Kickxellomycotina, Mortierellomycotina and Mucoromycotina
Type: C=Culture; F=Fruiting Bodies
Provider: CCBAS: The samples were provided by the Culture Collection of Basidiomycetes, Institute of Microbiology AS CR
CCF: The samples were provided by the Culture Collection of Fungi, Department of Botany, Faculty of Science of the Charles University in Prague
CFGM: The samples from fresh fruiting bodies collected throughout mycological excursions or from collections of Laboratory of Fungal Genetics and Metabolism, Institute of Microbiology AS CR
CMT: The samples from collections of Michal Tomšovský from Department of Forest Protection and Wildlife Management at Faculty of Forestry and Wood Technology, Mendel University in Brno
CJB: The samples from collections of Jan Borovička from Department of Environmental Geology and Geochemistry, Institute of Geology AS CR
PRM: The samples from the herbarium of National Museum in Prague
Ticks mark items that were included in overlapped analysis for all markers and or in single gene analyses; the column ITS1, ITS2, *EF-1α_A* and *RPB2_B*

Species	No.	Family	Order	Phyl. (Subphyl)	Type	Deposit Number	Provider	Overlap	ITS1	ITS2	*EF-1α_A*	*RPB2_B*
Abortiporus biennis	2	Coriolaceae	Polyporales	Bas	C	CCBAS498	CCBAS	✓	✓	✓	✓	✓
Abortiporus biennis	2	Coriolaceae	Polyporales	Bas	C	CCBAS498	CCBAS	✓	✓	✓	✓	✓
Absidia glauca	1	Cunninghamellaceae	Mucorales	Muc	C	42	CFGM	✓	✓	✓	✓	✓
Absidia sp.	1	Cunninghamellaceae	Mucorales	Muc	C	255	CFGM	✓	✓	✓	✓	✓
Acremonium variecolor	2	mitosporicHypocreales	Hypocreales	Asc	C	CCBAS497	CCBAS	✓	✓	✓	✓	✓
Acremonium variecolor	2	mitosporicHypocreales	Hypocreales	Asc	C	CCBAS832	CCBAS	✓	✓	✓	✓	✓
Agaricus altipes	1	Agaricaceae	Agaricales	Bas	F	A40	CJB	✓	✓	✓	✓	✓
Agaricus arvensis	1	Agaricaceae	Agaricales	Bas	F	A48	CJB	–	–	–	✓	✓

Agaricus augustus	1	Agaricaceae	Agaricales	Bas	C	CCBAS308	CCBAS	✓	✓	✓	✓	✓
Agaricus bisporus	2	Agaricaceae	Agaricales	Bas	C	CCBAS306	CCBAS	✓	✓	✓	✓	✓
Agaricus bisporus	2	Agaricaceae	Agaricales	Bas	C	CCBAS306	CCBAS	✓	✓	✓	✓	✓
Agaricus bitorquis	1	Agaricaceae	Agaricales	Bas	F	A4	CJB	✓	✓	✓	✓	✓
Agaricus cappellii	1	Agaricaceae	Agaricales	Bas	F	A51	CJB	–	–	–	✓	–
Agaricus chionodermus	2	Agaricaceae	Agaricales	Bas	F	A49	CJB	–	–	✓	✓	✓
Agaricus chionodermus	2	Agaricaceae	Agaricales	Bas	F	A41	CJB	✓	✓	✓	✓	✓
Agaricus flocculosipes	1	Agaricaceae	Agaricales	Bas	C	CCBAS308	CCBAS	✓	✓	✓	✓	✓
Agaricus julius	1	Agaricaceae	Agaricales	Bas	F	A42	CJB	✓	✓	✓	✓	✓
Agaricus kotlabae	1	Agaricaceae	Agaricales	Bas	F	A5	CJB	✓	✓	✓	✓	✓
Agaricus osecanus	1	Agaricaceae	Agaricales	Bas	F	A50	CJB	–	–	–	–	–
Agaricus pseudoumbrella	1	Agaricaceae	Agaricales	Bas	F	A52	CJB	–	✓	✓	✓	–
Agaricus sp.1	1	Agaricaceae	Agaricales	Bas	C	CCBAS301	CCBAS	✓	✓	✓	✓	✓
Agaricus sp.2	1	Agaricaceae	Agaricales	Bas	C	CCBAS301	CCBAS	✓	✓	✓	✓	✓
Agaricus xanthodermus	2	Agaricaceae	Agaricales	Bas	F	A3	CJB	✓	✓	✓	✓	✓
Agaricus xanthodermus	2	Agaricaceae	Agaricales	Bas	F	TZ11	CFGM	✓	✓	✓	✓	✓
Agrocybe aegerita	5	Bolbitiaceae	Agaricales	Bas	C	CCBAS753	CCBAS	✓	✓	✓	✓	✓
Agrocybe aegerita	5	Bolbitiaceae	Agaricales	Bas	C	CCBAS305	CCBAS	✓	✓	✓	✓	✓
Agrocybe aegerita	5	Bolbitiaceae	Agaricales	Bas	C	CCBAS303	CCBAS	✓	✓	✓	✓	✓
Agrocybe aegerita	5	Bolbitiaceae	Agaricales	Bas	C	CCBAS312	CCBAS	✓	✓	✓	✓	✓
Agrocybe aegerita	5	Bolbitiaceae	Agaricales	Bas	C	CCBAS303	CCBAS	✓	✓	✓	✓	✓
Agrocybe erebia	1	Bolbitiaceae	Agaricales	Bas	C	CCBAS291	CCBAS	✓	✓	✓	✓	✓
Agrocybe praecox	1	Bolbitiaceae	Agaricales	Bas	C	CCBAS641	CCBAS	✓	✓	✓	✓	✓
Agrocybe smithii	1	Bolbitiaceae	Agaricales	Bas	C	CCBAS642	CCBAS	✓	✓	✓	✓	✓
Agrocybe sp.	1	Bolbitiaceae	Agaricales	Bas	F	h37	CFGM	–	✓	✓	✓	–
Alternaria sp.	1	Pleosporaceae	Pleosporales	Asc	C	H02	CFGM	✓	✓	✓	✓	✓
Amanita caesarea	1	Amanitaceae	Agaricales	Bas	F	A45	CJB	✓	✓	✓	✓	✓
Amanita codinae	1	Amanitaceae	Agaricales	Bas	F	A19	CJB	–	✓	✓	✓	–
Amanita friabilis	2	Amanitaceae	Agaricales	Bas	F	A43	CJB	✓	✓	✓	✓	✓
Amanita friabilis	2	Amanitaceae	Agaricales	Bas	F	A23	CJB	✓	✓	✓	✓	✓

Amanita gilbertii	2	Amanitaceae	Agaricales	Bas	F	A15	CJB	✓	✓	✓	✓	✓
Amanita gilbertii	2	Amanitaceae	Agaricales	Bas	F	A22	CJB	✓	✓	✓	✓	✓
Amanita gracilior	2	Amanitaceae	Agaricales	Bas	F	A17	CJB	✓	✓	✓	✓	✓
Amanita gracilior	2	Amanitaceae	Agaricales	Bas	F	A34	CJB	✓	✓	✓	✓	✓
Amanita groenlandica	1	Amanitaceae	Agaricales	Bas	F	A47	CJB	✓	✓	✓	✓	✓
Amanita inopinata	1	Amanitaceae	Agaricales	Bas	F	A37	CJB	–	–	✓	✓	✓
Amanita muscaria	3	Amanitaceae	Agaricales	Bas	F	h13	CFGM	✓	✓	✓	✓	✓
Amanita muscaria	3	Amanitaceae	Agaricales	Bas	F	h14	CFGM	✓	✓	✓	✓	✓
Amanita muscaria	3	Amanitaceae	Agaricales	Bas	F	A16	CJB	✓	✓	✓	✓	✓
Amanita pachyvolvata	1	Amanitaceae	Agaricales	Bas	F	A33	CJB	✓	✓	✓	✓	✓
Amanita phalloides	1	Amanitaceae	Agaricales	Bas	F	TZ2	CFGM	✓	✓	✓	✓	✓
Amanita regalis	1	Amanitaceae	Agaricales	Bas	F	A46	CJB	✓	✓	✓	✓	✓
Amanita singeri	1	Amanitaceae	Agaricales	Bas	F	A20	CJB	–	✓	✓	✓	–
Amanita solitaria	2	Amanitaceae	Agaricales	Bas	F	A2	CJB	–	✓	✓	✓	–
Amanita solitaria	2	Amanitaceae	Agaricales	Bas	F	A12	CJB	–	✓	✓	✓	–
Amanita sp.1	1	Amanitaceae	Agaricales	Bas	F	A25	CJB	✓	✓	✓	✓	✓
Amanita sp.2	1	Amanitaceae	Agaricales	Bas	F	A27	CJB	–	–	✓	✓	✓
Amanita sp.3	1	Amanitaceae	Agaricales	Bas	F	A28	CJB	–	✓	✓	–	–
Amanita sp.4	1	Amanitaceae	Agaricales	Bas	F	A29	CJB	–	–	✓	✓	–
Amanita sp.5	1	Amanitaceae	Agaricales	Bas	F	A30	CJB	–	✓	✓	✓	–
Amanita sp.6	1	Amanitaceae	Agaricales	Bas	F	A31	CJB	–	✓	✓	✓	–
Amanita sp.7	1	Amanitaceae	Agaricales	Bas	F	A29	CJB	✓	✓	✓	✓	✓
Amanita strobiliformis	1	Amanitaceae	Agaricales	Bas	F	A13	CJB	–	✓	✓	✓	–
Amanita submembranacea	1	Amanitaceae	Agaricales	Bas	F	A24	CJB	✓	✓	✓	✓	✓
Amanita vaginata	3	Amanitaceae	Agaricales	Bas	F	A32	CJB	–	✓	✓	✓	–
Amanita vaginata	3	Amanitaceae	Agaricales	Bas	F	A35	CJB	–	✓	✓	–	–
Amanita vaginata	3	Amanitaceae	Agaricales	Bas	F	A36	CJB	–	–	✓	✓	✓
Amanita verna	1	Amanitaceae	Agaricales	Bas	F	A44	CJB	✓	✓	✓	✓	✓
Amanita vittadinii	1	Amanitaceae	Agaricales	Bas	F	A1	CJB	–	✓	✓	✓	–
Amanitopsis sp.1	1	Amanitaceae	Agaricales	Bas	F	A6	CJB	–	✓	✓	✓	–

Amanitopsis sp.2	1	Amanitaceae	Agaricales	Bas	F	A7	CJB	–	✓	✓	✓	–
Amanitopsis sp.3	1	Amanitaceae	Agaricales	Bas	F	A8	CJB	–	✓	✓	✓	–
Amanitopsis sp.4	1	Amanitaceae	Agaricales	Bas	F	A9	CJB	✓	✓	✓	✓	✓
Amanitopsis sp.5	1	Amanitaceae	Agaricales	Bas	F	A10	CJB	✓	✓	✓	✓	✓
Amanitopsis sp.6	1	Amanitaceae	Agaricales	Bas	F	A11	CJB	✓	✓	✓	✓	✓
Amyloporia sp.	1	Polyporaceae	Polyporales	Bas	C	CCBAS668	CCBAS	–	–	✓	✓	–
Antrodia heteromorpha	1	Fomitopsidaceae	Polyporales	Bas	C	CCBAS105	CCBAS	–	–	✓	–	–
Armillaria calvescens	1	Physalacriaceae	Agaricales	Bas	C	CCBAS330	CCBAS	✓	✓	✓	✓	✓
Armillaria cepistipes	2	Physalacriaceae	Agaricales	Bas	F	506	CMT	✓	✓	✓	✓	✓
Armillaria cepistipes	2	Physalacriaceae	Agaricales	Bas	F	513	CMT	✓	✓	✓	✓	✓
Armillaria gemina	1	Physalacriaceae	Agaricales	Bas	C	CCBAS326	CCBAS	✓	✓	✓	✓	✓
Armillaria mellea	2	Physalacriaceae	Agaricales	Bas	F	574	CMT	✓	✓	✓	✓	✓
Armillaria mellea	2	Physalacriaceae	Agaricales	Bas	F	588	CMT	✓	✓	✓	✓	✓
Armillaria ostoyae	4	Physalacriaceae	Agaricales	Bas	F	560	CMT	✓	✓	✓	✓	✓
Armillaria ostoyae	4	Physalacriaceae	Agaricales	Bas	C	CCBAS325	CCBAS	✓	✓	✓	✓	✓
Armillaria ostoyae	4	Physalacriaceae	Agaricales	Bas	C	CCBAS325	CCBAS	✓	✓	✓	✓	✓
Armillaria ostoyae	4	Physalacriaceae	Agaricales	Bas	F	504	CMT	✓	✓	✓	✓	✓
Aspergillus penicillioides	1	Trichocomaceae	Eurotiales	Asc	C	MK2001	CFGM	✓	✓	✓	✓	✓
Asterophora lycoperdoides	1	Lyophyllaceae	Agaricales	Bas	F	TZ5	CFGM	–	✓	✓	✓	–
Athelia bombacina	1	Atheliaceae	Atheliales	Bas	C	B1/1	CFGM	–	✓	✓	✓	–
Baeospora myosura	1	Tricholomataceae	Agaricales	Bas	F	h55	CFGM	✓	✓	✓	✓	✓
Beauveria felina	1	Cordycipitaceae	Hypocreales	Asc	C	AN1/09	CFGM	✓	✓	✓	✓	✓
Bettsia fastidia	1	IncertaeSedis	IncertaeSedis	Asc	C	CCF3693	CCF	–	✓	✓	–	–
Biscogniauxia nummularia	2	Xylariaceae	Xylariales	Asc	C	H07	CFGM	✓	✓	✓	✓	✓
Biscogniauxia nummularia	2	Xylariaceae	Xylariales	Asc	C	H08	CFGM	✓	✓	✓	✓	✓
Bjerkandera adusta	3	Meruliaceae	Polyporales	Bas	C	CCBAS749	CCBAS	✓	✓	✓	✓	✓
Bjerkandera adusta	3	Meruliaceae	Polyporales	Bas	C	CCBAS230	CCBAS	✓	✓	✓	✓	✓
Bjerkandera adusta	3	Meruliaceae	Polyporales	Bas	C	CCBAS244	CCBAS	✓	✓	✓	✓	✓
Boletus chrysoxanthus	1	Boletaceae	Boletales	Bas	F	B3	CJB	–	✓	✓	✓	–
Boletus edulis	2	Boletaceae	Boletales	Bas	F	TZ12	CFGM	✓	✓	✓	✓	✓

Boletus edulis	2	Boletaceae	Boletales	Bas	F	h47	CFGM	✓	✓	✓	✓	✓
Boletus legaliae	1	Boletaceae	Boletales	Bas	F	B2	CJB	✓	✓	✓	✓	✓
Boletus pruinatus	1	Boletaceae	Boletales	Bas	F	h2	CFGM	✓	✓	✓	✓	✓
Boletus spinari	1	Boletaceae	Boletales	Bas	F	B1	CJB	–	✓	✓	✓	–
Calcarisporium arbuscula	1	IncertaeSedis	Hypocreales	Asc	C	CCBAS285	CCBAS	✓	✓	✓	✓	✓
Camarosporium brabeji	1	mitosporicPleosporales	Pleosporales	Asc	C	H21	CFGM	✓	✓	✓	✓	✓
Ceriporia camaresiana	1	Phanerochaetaceae	Polyporales	Bas	C	CCBAS558	CCBAS	–	✓	✓	✓	–
Chalciporus piperatus	1	Boletaceae	Boletales	Bas	F	h27	CFGM	✓	✓	✓	✓	✓
Chrysosporium chiropterorum	1	mitosporicOnygenales	Onygenales	Asc	C	45495	CFGM	–	✓	✓	✓	–
Chrysosporium speluncarum	1	mitosporicOnygenales	Onygenales	Asc	C	CCF3760	CCF	–	✓	✓	✓	–
Chrysosporium sulphureum	1	mitosporicOnygenales	Onygenales	Asc	C	AK16/12	CFGM	✓	✓	✓	✓	✓
Chrysosporium undulatum	1	mitosporicOnygenales	Onygenales	Asc	C	BE1	CFGM	✓	✓	✓	✓	✓
Circinella lacrymispora	1	Cunninghamellaceae	Mucorales	Muc	C	CBS101.757	CFGM	✓	✓	✓	✓	✓
Cladophialophora sp.	1	Herpotrichiellaceae	Chaetothyriales	Asc	C	V475	CFGM	–	–	–	✓	✓
Cladosporium herbarum	1	Cladosporiaceae	Capnodiales	Asc	C	R191	CFGM	✓	✓	✓	✓	✓
Cladosporium sp.	1	Cladosporiaceae	Capnodiales	Asc	C	R973	CFGM	✓	✓	✓	✓	✓
Clavulicium globosum	1	Clavulinaceae	Cantharellales	Bas	C	CCBAS653	CCBAS	✓	✓	✓	✓	✓
Clitocybe subditopoda	1	Tricholomataceae	Agaricales	Bas	F	h46	CFGM	–	✓	✓	✓	–
Clitocybula familia	5	Tricholomataceae	Agaricales	Bas	F	X14	CJB	✓	✓	✓	✓	✓
Clitocybula familia	5	Tricholomataceae	Agaricales	Bas	F	X21	CJB	✓	✓	✓	✓	✓
Clitocybula familia	5	Tricholomataceae	Agaricales	Bas	F	X2	CJB	✓	✓	✓	✓	✓
Clitocybula familia	5	Tricholomataceae	Agaricales	Bas	F	X13	CJB	✓	✓	✓	✓	✓
Clitocybula familia	5	Tricholomataceae	Agaricales	Bas	F	X15	CJB	✓	✓	✓	✓	✓
Clitocybula sp.1	1	Tricholomataceae	Agaricales	Bas	F	X16	CJB	✓	✓	✓	✓	✓
Clitocybula sp.2	1	Tricholomataceae	Agaricales	Bas	F	X62	CJB	✓	✓	✓	✓	✓
Clitocybula taniae	1	Tricholomataceae	Agaricales	Bas	F	X61	CJB	✓	✓	✓	✓	✓
Clitopilus passeckerianus	1	Entolomataceae	Agaricales	Bas	C	CCBAS738	CCBAS	–	✓	✓	✓	–
Clitopilus prunulus	1	Entolomataceae	Agaricales	Bas	F	TZ30	CFGM	✓	✓	✓	✓	✓
Clonostachys rosea	1	Bionectriaceae	Hypocreales	Asc	C	174	CFGM	✓	✓	✓	✓	✓
Colletotrichum coccodes	1	Glomerellaceae	Glomerellales	Asc	C	AK273/05	CFGM	✓	✓	✓	✓	✓

Coniochaeta ligniaria	1	Coniochaetaceae	Coniochaetales	Asc	C	AK213/05	CFGM	–	✓	✓	✓	–
Coprinellus bisporus	4	Psathyrellaceae	Agaricales	Bas	C	CCBAS359	CCBAS	✓	✓	✓	✓	✓
Coprinellus bisporus	4	Psathyrellaceae	Agaricales	Bas	C	CCBAS639	CCBAS	–	–	✓	✓	✓
Coprinellus bisporus	4	Psathyrellaceae	Agaricales	Bas	C	CCBAS356	CCBAS	✓	✓	✓	✓	✓
Coprinellus bisporus	4	Psathyrellaceae	Agaricales	Bas	C	CCBAS356	CCBAS	✓	✓	✓	✓	✓
Coprinellus sp.	1	Psathyrellaceae	Agaricales	Bas	C	CCF3733	CCF	✓	✓	✓	✓	✓
Cordyceps confragosa	6	Clavicipitaceae	Hypocreales	Asc	C	CCBAS313	CCBAS	✓	✓	✓	✓	✓
Cordyceps confragosa	6	Clavicipitaceae	Hypocreales	Asc	C	CCBAS283	CCBAS	–	–	✓	✓	✓
Cordyceps confragosa	6	Clavicipitaceae	Hypocreales	Asc	C	CCBAS777	CCBAS	✓	✓	✓	✓	✓
Cordyceps confragosa	6	Clavicipitaceae	Hypocreales	Asc	C	CCBAS280	CCBAS	✓	✓	✓	✓	✓
Cordyceps confragosa	6	Clavicipitaceae	Hypocreales	Asc	C	CCBAS626	CCBAS	✓	✓	✓	✓	✓
Cordyceps confragosa	6	Clavicipitaceae	Hypocreales	Asc	C	CCBAS541	CCBAS	–	–	✓	✓	✓
Cordyceps sp.1	1	Clavicipitaceae	Hypocreales	Asc	C	AK56_08	CFGM	✓	✓	✓	✓	✓
Cordyceps sp.2	1	Clavicipitaceae	Hypocreales	Asc	C	AK50/08	CFGM	✓	✓	✓	✓	✓
Cortinarius alnetorum	3	Cortinariaceae	Agaricales	Bas	F	X23	CJB	✓	✓	✓	✓	✓
Cortinarius alnetorum	3	Cortinariaceae	Agaricales	Bas	F	X24	CJB	✓	✓	✓	✓	✓
Cortinarius alnetorum	3	Cortinariaceae	Agaricales	Bas	F	X25	CJB	✓	✓	✓	✓	✓
Cortinarius bibulus	1	Cortinariaceae	Agaricales	Bas	F	X28	CJB	–	✓	✓	✓	–
Cortinarius glaucopus	1	Cortinariaceae	Agaricales	Bas	F	h7	CFGM	✓	✓	✓	✓	✓
Cortinarius pardinus	2	Cortinariaceae	Agaricales	Bas	F	X11	CJB	✓	✓	✓	✓	✓
Cortinarius pardinus	2	Cortinariaceae	Agaricales	Bas	F	X43	CJB	✓	✓	✓	✓	✓
Cortinarius pseudovulpinus	1	Cortinariaceae	Agaricales	Bas	F	X3	CJB	–	–	✓	–	✓
Cortinarius sp.	1	Cortinariaceae	Agaricales	Bas	F	X42	CJB	✓	✓	✓	✓	✓
Cortinarius variecolor	1	Cortinariaceae	Agaricales	Bas	F	X47	CJB	✓	✓	✓	✓	✓
Craterellus tubaeformis	1	Cantharellaceae	Cantharellales	Bas	F	h36	CFGM	✓	✓	✓	✓	✓
Crucibulum laeve	1	Nidulariaceae	Agaricales	Bas	F	TZ9	CFGM	✓	✓	✓	✓	✓
Cunninghamella echinulata	1	Cunninghamellaceae	Mucorales	Muc	C	276	CFGM	✓	✓	✓	✓	✓
Custingophora olivacea	1	IncertaeSedis	IncertaeSedis	Asc	C	CO	CFGM	✓	✓	✓	✓	✓
Cyathus striatus	2	Nidulariaceae	Agaricales	Bas	F	TZ8	CFGM	✓	✓	✓	✓	✓
Cyathus striatus	2	Nidulariaceae	Agaricales	Bas	C	CCBAS807	CCBAS	✓	✓	✓	✓	✓

Daedalea quercina	1	Coriolaceae	Polyporales	Bas	C	CCBAS529	CCBAS	–	✓	✓	✓	–
Daedaleopsis confragosa	1	Coriolaceae	Polyporales	Bas	C	CCBAS530	CCBAS	✓	✓	✓	✓	✓
Daedaleopsis tricolor	1	Polyporaceae	Polyporales	Bas	C	CCBAS590	CCBAS	–	–	–	✓	✓
Daldinia childiae	1	Xylariaceae	Xylariales	Asc	C	H14	CFGM	✓	✓	✓	✓	✓
Dermocybe cinnamomeolutea	1	Cortinariaceae	Agaricales	Bas	F	h61	CFGM	✓	✓	✓	✓	✓
Dermocybe semisanguinea	1	Cortinariaceae	Agaricales	Bas	F	h38	CFGM	✓	✓	✓	✓	✓
Desmazierella acicola	1	Sarcoscyphaceae	Pezizales	Asc	C	H20	CFGM	✓	✓	✓	✓	✓
Diaporthe oncostoma	1	Valsaceae	Diaporthales	Asc	C	AK199/05	CFGM	✓	✓	✓	✓	✓
Diaporthe sp.	1	Diaporthaceae	Diaporthales	Asc	C	H18	CFGM	✓	✓	✓	✓	✓
Diatrypella pulvinata	1	Diatrypaceae	Xylariales	Asc	C	H48	CFGM	✓	✓	✓	✓	✓
Dichomitus squalens	1	Polyporaceae	Polyporales	Bas	C	CCBAS751	CCBAS	✓	✓	✓	✓	✓
Dimargaris bacillispora	1	Dimargaritaceae	Dimargaritales	Kic	C	DIG3	CFGM	✓	✓	✓	✓	✓
Dothideomycetes sp.	1	Dothideomycetes sp.	Dothideomycetes sp.	Asc	C	R956	CFGM	✓	✓	✓	✓	✓
Endoptychum depressum	2	Agaricaceae	Agaricales	Bas	C	CCBAS302	CCBAS	✓	✓	✓	✓	✓
Endoptychum depressum	2	Agaricaceae	Agaricales	Bas	C	CCBAS302	CCBAS	✓	✓	✓	✓	✓
Entocybe turbidum	1	Entolomataceae	Agaricales	Bas	F	X6	CJB	–	✓	✓	✓	–
Entoloma lividum	1	Entolomataceae	Agaricales	Bas	F	X7	CJB	✓	✓	✓	✓	✓
Entoloma rhodopolium	1	Entolomataceae	Agaricales	Bas	F	X9	CJB	–	✓	✓	–	–
Entoloma sp.	1	Entolomataceae	Agaricales	Bas	F	X12	CJB	–	–	✓	✓	–
Entyloma microsporum	1	Entylomataceae	Entylomatales	Bas	C	CCBAS398	CCBAS	–	–	✓	✓	–
Epichloe typhina	3	Clavicipitaceae	Hypocreales	Asc	C	CCBAS341	CCBAS	✓	✓	✓	✓	✓
Epichloe typhina	3	Clavicipitaceae	Hypocreales	Asc	C	CCBAS632	CCBAS	✓	✓	✓	✓	✓
Epichloe typhina	3	Clavicipitaceae	Hypocreales	Asc	C	CCBAS341	CCBAS	✓	✓	✓	✓	✓
Fayodia gracilipes	1	Tricholomataceae	Agaricales	Bas	C	CCBAS805	CCBAS	✓	✓	✓	✓	✓
Fistulina hepatica	1	Fistulinaceae	Agaricales	Bas	C	CCBAS532	CCBAS	✓	✓	✓	✓	✓
Flammulina elastica	1	Physalacriaceae	Agaricales	Bas	C	CCBAS363	CCBAS	✓	✓	✓	✓	✓
Fusarium solani	3	Nectriaceae	Hypocreales	Asc	C	CCBASK95	CCBAS	✓	✓	✓	✓	✓
Fusarium solani	3	Nectriaceae	Hypocreales	Asc	C	CCBAS844	CCBAS	✓	✓	✓	✓	✓
Fusarium solani	3	Nectriaceae	Hypocreales	Asc	C	CCBASK95	CCBAS	✓	✓	✓	✓	✓
Fusarium sp.	1	Nectriaceae	Hypocreales	Asc	C	CCBAS850	CCBAS	✓	✓	✓	✓	✓

Fuscoporia torulosa	1	Hymenochaetaceae	Hymenochaetales	Bas	C	CCBAS759	CCBAS	✓	✓	✓	✓	✓
Fusicladium sp.1	1	Sympoventuriaceae	Venturiales	Asc	C	AK80/06	CFGM	–	✓	✓	✓	–
Fusicladium sp.2	1	Sympoventuriaceae	Venturiales	Asc	C	V432	CFGM	–	✓	✓	✓	–
Galactomyces geotrichum	1	Dipodascaceae	Saccharomycetales	Asc	C	409a	CFGM	✓	✓	✓	✓	✓
Galerina marginata	1	Strophariaceae	Agaricales	Bas	F	X10	CJB	✓	✓	✓	✓	✓
Ganoderma applanatum	2	Ganodermataceae	Polyporales	Bas	C	CCBAS707	CCBAS	✓	✓	✓	✓	✓
Ganoderma applanatum	2	Ganodermataceae	Polyporales	Bas	C	CCBAS746	CCBAS	✓	✓	✓	✓	✓
Ganoderma australe	1	Ganodermataceae	Polyporales	Bas	C	CCBAS744	CCBAS	✓	✓	✓	✓	✓
Geomyces sp.	1	Pseudeurotiaceae	IncertaeSedis	Asc	C	AK77/11	CFGM	✓	✓	✓	✓	✓
Gomphidius glutinosus	1	Gomphidiaceae	Boletales	Bas	F	h11	CFGM	–	–	✓	✓	–
Gongronella lacrispora	2	Cunninghamellaceae	Mucorales	Muc	C	CBS244.62	CFGM	✓	✓	✓	✓	✓
Gongronella lacrispora	2	Cunninghamellaceae	Mucorales	Muc	C	CCF3867	CCF	✓	✓	✓	✓	✓
Graphium fimbriisporum	1	Microascaceae	Microascales	Asc	C	H11	CFGM	✓	✓	✓	✓	✓
Graphium penicillioides	1	Graphiaceae	Microascales	Asc	C	H46	CFGM	–	✓	✓	✓	–
Graphium sp.	1	Microascaceae	Microascales	Asc	C	CCF3566	CCF	✓	✓	✓	✓	✓
Guignardia vaccinii	1	Phyllostictaceae	Botryosphaeriales	Asc	C	CCBAS931	CCBAS	✓	✓	✓	✓	✓
Gymnopilus hybridus	1	Cortinariaceae	Agaricales	Bas	C	CCBAS287	CCBAS	–	–	✓	✓	✓
Gymnopilus penetrans	1	Cortinariaceae	Agaricales	Bas	F	TZ1	CFGM	–	✓	–	✓	✓
Gymnopilus sp.	1	Cortinariaceae	Agaricales	Bas	F	TZ29	CFGM	✓	✓	✓	✓	✓
Gymnopus aquosus	1	Tricholomataceae	Agaricales	Bas	F	G30	CMT	✓	✓	✓	✓	✓
Gymnopus dryophilus	3	Tricholomataceae	Agaricales	Bas	F	G98	CMT	✓	✓	✓	✓	✓
Gymnopus dryophilus	3	Tricholomataceae	Agaricales	Bas	F	G61	CMT	✓	✓	✓	✓	✓
Gymnopus dryophilus	3	Tricholomataceae	Agaricales	Bas	F	G31	CMT	✓	✓	✓	✓	✓
Gymnopus ocior	1	Tricholomataceae	Agaricales	Bas	C	CCBAS281	CCBAS	✓	✓	✓	✓	✓
Gymnopus sp.	1	Tricholomataceae	Agaricales	Bas	F	TZ17	CFGM	–	✓	✓	✓	–
Hebeloma bulbiferum	2	Cortinariaceae	Agaricales	Bas	F	X26	CJB	✓	✓	✓	✓	✓
Hebeloma bulbiferum	2	Cortinariaceae	Agaricales	Bas	F	X50	CJB	✓	✓	✓	✓	✓
Hebeloma cavipes	2	Cortinariaceae	Agaricales	Bas	F	X27	CJB	✓	✓	✓	✓	✓
Hebeloma cavipes	2	Cortinariaceae	Agaricales	Bas	F	X18	CJB	✓	✓	✓	✓	✓
Hebeloma crustuliniforme	1	Cortinariaceae	Agaricales	Bas	F	X20	CJB	✓	✓	✓	✓	✓

Hebeloma laterinum	1	Cortinariaceae	Agaricales	Bas	F	X34	CJB	✓	✓	✓	✓	✓
Hebeloma mesophaeum	3	Cortinariaceae	Agaricales	Bas	F	X31	CJB	✓	✓	✓	✓	✓
Hebeloma mesophaeum	3	Cortinariaceae	Agaricales	Bas	F	X32	CJB	✓	✓	✓	✓	✓
Hebeloma mesophaeum	3	Cortinariaceae	Agaricales	Bas	F	X33	CJB	✓	✓	✓	✓	✓
Hebeloma sacchariolens	2	Cortinariaceae	Agaricales	Bas	F	X56	CJB	–	–	✓	✓	✓
Hebeloma sacchariolens	2	Cortinariaceae	Agaricales	Bas	F	X30	CJB	✓	✓	✓	✓	✓
Hebeloma sordidum	6	Cortinariaceae	Agaricales	Bas	F	X51	CJB	✓	✓	✓	✓	✓
Hebeloma sordidum	6	Cortinariaceae	Agaricales	Bas	F	X48	CJB	✓	✓	✓	✓	✓
Hebeloma sordidum	6	Cortinariaceae	Agaricales	Bas	F	X58	CJB	✓	✓	✓	✓	✓
Hebeloma sordidum	6	Cortinariaceae	Agaricales	Bas	F	X52	CJB	✓	✓	✓	✓	✓
Hebeloma sordidum	6	Cortinariaceae	Agaricales	Bas	F	X53	CJB	✓	✓	✓	✓	✓
Hebeloma sordidum	6	Cortinariaceae	Agaricales	Bas	F	X60	CJB	✓	✓	✓	✓	✓
Hebeloma sp.1	1	Cortinariaceae	Agaricales	Bas	C	CCBAS834	CCBAS	✓	✓	✓	✓	✓
Hebeloma sp.2	1	Cortinariaceae	Agaricales	Bas	F	X19	CJB	✓	✓	✓	✓	✓
Hebeloma sp.3	1	Cortinariaceae	Agaricales	Bas	F	X55	CJB	✓	✓	✓	✓	✓
Helotiales sp.	1	IncertaeSedis	Helotiales	Asc	C	VK283	CFGM	–	✓	✓	✓	–
Hericium coralloides	1	Hericiaceae	Russulales	Bas	C	CCBAS548	CCBAS	✓	✓	✓	✓	✓
Hericium erinaceus	1	Hericiaceae	Russulales	Bas	C	CCBAS551	CCBAS	✓	✓	✓	✓	✓
Heterobasidion abietinum	3	Bondarzewiaceae	Russulales	Bas	F	36H	CMT	✓	✓	✓	✓	✓
Heterobasidion abietinum	3	Bondarzewiaceae	Russulales	Bas	F	35H	CMT	✓	✓	✓	✓	✓
Heterobasidion abietinum	3	Bondarzewiaceae	Russulales	Bas	F	100H	CMT	✓	✓	✓	✓	✓
Heterobasidion annosum	3	Bondarzewiaceae	Russulales	Bas	F	47H	CMT	✓	✓	✓	✓	✓
Heterobasidion annosum	3	Bondarzewiaceae	Russulales	Bas	F	87H	CMT	✓	✓	✓	✓	✓
Heterobasidion annosum	3	Bondarzewiaceae	Russulales	Bas	F	89H	CMT	✓	✓	✓	✓	✓
Heterobasidion parviporum	3	Bondarzewiaceae	Russulales	Bas	F	18_10H	CMT	✓	✓	✓	✓	✓
Heterobasidion parviporum	3	Bondarzewiaceae	Russulales	Bas	F	111H	CMT	✓	✓	✓	✓	✓
Heterobasidion parviporum	3	Bondarzewiaceae	Russulales	Bas	F	93H	CMT	✓	✓	✓	✓	✓
Hohenbuehelia auriscalpium	1	Pleurotaceae	Agaricales	Bas	C	CCBAS373	CCBAS	✓	✓	✓	✓	✓
Hydnopolyporus fimbriatus	1	Meripilaceae	Polyporales	Bas	C	CCBAS615	CCBAS	✓	✓	✓	✓	✓
Hygrophoropsis aurantiaca	2	Hygrophoropsidaceae	Boletales	Bas	F	h23	CFGM	✓	✓	✓	✓	✓

Hygrophoropsis aurantiaca	2	Hygrophoropsidaceae	Boletales	Bas	F	TZ28	CFGM	✓	✓	✓	✓	✓
Hygrophorus hedrychii	1	Hygrophoraceae	Agaricales	Bas	F	X38	CJB	–	–	✓	–	–
Hygrophorus persicolor	1	Hygrophoraceae	Agaricales	Bas	F	X39	CJB	–	–	–	–	–
Hymenopellis	1	Physalacriaceae	Agaricales	Bas	C	CCBAS426	CCBAS	–	✓	✓	✓	–
Hymenopellis radicata	5	Physalacriaceae	Agaricales	Bas	C	CCBAS442	CCBAS	✓	✓	✓	✓	✓
Hymenopellis radicata	5	Physalacriaceae	Agaricales	Bas	C	CCBAS444	CCBAS	✓	✓	✓	✓	✓
Hymenopellis radicata	5	Physalacriaceae	Agaricales	Bas	C	CCBAS713	CCBAS	✓	✓	✓	✓	✓
Hymenopellis radicata	5	Physalacriaceae	Agaricales	Bas	C	CCBAS711	CCBAS	✓	✓	✓	✓	✓
Hymenopellis radicata	5	Physalacriaceae	Agaricales	Bas	C	CCBAS711	CCBAS	✓	✓	✓	✓	✓
Hyphoderma definitum	1	Corticiaceae	Corticiales	Bas	C	CBS103.982	CFGM	–	–	✓	✓	✓
Hyphoderma guttuliferum	1	Corticiaceae	Corticiales	Bas	C	CBS107.303	CFGM	✓	✓	✓	✓	✓
Hyphoderma orphanellum	1	Corticiaceae	Corticiales	Bas	C	CBS105814	CFGM	✓	✓	✓	✓	✓
Hyphoderma sp.	1	Corticiaceae	Corticiales	Bas	C	90	CFGM	✓	✓	✓	✓	✓
Hyphodontia rimosissima	1	Corticiaceae	Corticiales	Bas	C	CBS105045	CFGM	✓	✓	✓	✓	✓
Hypholoma fasciculare	2	Strophariaceae	Agaricales	Bas	C	CCBAS358	CCBAS	✓	✓	✓	✓	✓
Hypholoma fasciculare	2	Strophariaceae	Agaricales	Bas	C	CCBAS381	CCBAS	✓	✓	✓	✓	✓
Hypholoma marginatum	1	Strophariaceae	Agaricales	Bas	F	P33	CJB	–	✓	✓	–	✓
Hypoxylon howeianum	1	Xylariaceae	Xylariales	Asc	C	H29	CFGM	✓	✓	✓	✓	✓
Ilyonectria radicicola	1	Nectriaceae	Hypocreales	Asc	C	S1	CFGM	✓	✓	✓	✓	✓
Infundibulicybe gibba	1	Tricholomataceae	Agaricales	Bas	F	TZ16	CFGM	✓	✓	✓	✓	✓
Inocutis dryophilus	1	Hymenochaetaceae	Hymenochaetales	Bas	C	CCBAS703	CCBAS	✓	✓	✓	✓	✓
Inocybe dulcamara	1	Inocybaceae	Agaricales	Bas	F	X1	CJB	✓	✓	✓	✓	✓
Inocybe napipes	1	Inocybaceae	Agaricales	Bas	F	X36	CJB	✓	✓	✓	✓	✓
Inocybe subcarpta	1	Inocybaceae	Agaricales	Bas	F	X37	CJB	✓	✓	✓	✓	✓
Inonotus andersonii	2	Hymenochaetaceae	Hymenochaetales	Bas	F	PH76	CMT	✓	✓	✓	✓	✓
Inonotus andersonii	2	Hymenochaetaceae	Hymenochaetales	Bas	F	1007	CMT	✓	✓	✓	✓	✓
Inonotus nidus-pici	1	Hymenochaetaceae	Hymenochaetales	Bas	F	PH75	CMT	✓	✓	✓	✓	✓
Inonotus obliquus	4	Hymenochaetaceae	Hymenochaetales	Bas	C	CCBAS559	CCBAS	✓	✓	✓	✓	✓
Inonotus obliquus	4	Hymenochaetaceae	Hymenochaetales	Bas	F	1001	CMT	✓	✓	✓	✓	✓
Inonotus obliquus	4	Hymenochaetaceae	Hymenochaetales	Bas	F	1096	CMT	✓	✓	✓	✓	✓

Inonotus obliquus	4	Hymenochaetaceae	Hymenochaetales	Bas	F	U5	CMT	✓	✓	✓	✓	✓
Irpex lacteus	2	IncertaeSedis	Polyporales	Bas	C	CCBAS369	CCBAS	✓	✓	✓	✓	✓
Irpex lacteus	2	IncertaeSedis	Polyporales	Bas	C	CCBAS196	CCBAS	✓	✓	✓	✓	✓
Irpex sp.	1	IncertaeSedis	Polyporales	Bas	C	CCBAS694	CCBAS	✓	✓	✓	✓	✓
Isaria farinosa	1	Cordycipitaceae	Hypocreales	Asc	C	CCBAS672	CCBAS	✓	✓	✓	✓	✓
Ischnoderma benzoinum	1	Fomitopsidaceae	Polyporales	Bas	C	CCBAS656	CCBAS	✓	✓	✓	✓	✓
Kretzschmaria deusta	1	Xylariaceae	Xylariales	Asc	C	AK220/05	CFGM	✓	✓	✓	✓	✓
Laccaria amethystina	1	Tricholomataceae	Agaricales	Bas	F	TZ26	CFGM	✓	✓	✓	✓	✓
Laccaria sp.	1	Hydnangiaceae	Agaricales	Bas	F	X45	CJB	✓	✓	✓	✓	✓
Lactarius deterrimus	1	Russulaceae	Russulales	Bas	F	TZ22	CFGM	✓	✓	✓	✓	✓
Lactarius rufus	1	Russulaceae	Russulales	Bas	F	h24	CFGM	✓	✓	✓	✓	✓
Laetiporus montanus	2	Coriolaceae	Polyporales	Bas	F	L20	CMT	✓	✓	✓	✓	✓
Laetiporus montanus	2	Coriolaceae	Polyporales	Bas	F	L12	CMT	✓	✓	✓	✓	✓
Laetiporus sulphureus	5	Coriolaceae	Polyporales	Bas	C	CCBAS563	CCBAS	✓	✓	✓	✓	✓
Laetiporus sulphureus	5	Coriolaceae	Polyporales	Bas	F	TZ13	CFGM	✓	✓	✓	✓	✓
Laetiporus sulphureus	5	Coriolaceae	Polyporales	Bas	F	ZB09	CMT	✓	✓	✓	✓	✓
Laetiporus sulphureus	5	Coriolaceae	Polyporales	Bas	F	ZB12	CMT	✓	✓	✓	✓	✓
Laetiporus sulphureus	5	Coriolaceae	Polyporales	Bas	F	ZB20	CMT	✓	✓	✓	✓	✓
Langermannia gigantea	1	Agaricaceae	Agaricales	Bas	C	CCBAS808	CCBAS	–	–	✓	✓	✓
Lecanicillium attenuatum	1	Clavicipitaceae	Hypocreales	Asc	C	CCBAS626	CCBAS	–	–	✓	✓	✓
Lecanicillium kalimantanense	1	Clavicipitaceae	Hypocreales	Asc	C	CCBAS522	CCBAS	✓	✓	✓	✓	✓
Lecanicillium sp.	1	Clavicipitaceae	Hypocreales	Asc	C	V289	CFGM	✓	✓	✓	✓	✓
Leccinum variicolor	1	Boletaceae	Boletales	Bas	F	h48	CFGM	✓	✓	✓	✓	✓
Leccinum versipelle	1	Boletaceae	Boletales	Bas	F	h52	CFGM	–	✓	✓	✓	–
Lentinula edodes	2	Omphalotaceae	Agaricales	Bas	C	CCBAS648	CCBAS	✓	✓	✓	✓	✓
Lentinula edodes	2	Omphalotaceae	Agaricales	Bas	C	CCBAS390	CCBAS	–	✓	✓	✓	–
Lentinus bertieri	1	Polyporaceae	Polyporales	Bas	C	CCBAS598	CCBAS	✓	✓	✓	✓	✓
Leotiales sp.	1	IncertaeSedis	Leotiales	Asc	C	AK91/11	CFGM	✓	✓	✓	✓	✓
Lepista irina	3	Tricholomataceae	Agaricales	Bas	C	CCBAS838	CCBAS	✓	✓	✓	✓	✓
Lepista irina	3	Tricholomataceae	Agaricales	Bas	C	CCBAS726	CCBAS	✓	✓	✓	✓	✓

Lepista irina	3	Tricholomataceae	Agaricales	Bas	C	CCBAS497	CCBAS	–	–	–	✓	✓
Lepista nuda	1	Tricholomataceae	Agaricales	Bas	C	CCBAS394	CCBAS	✓	✓	✓	✓	✓
Lepista sordida	2	Tricholomataceae	Agaricales	Bas	C	CCBAS761	CCBAS	✓	✓	✓	✓	✓
Lepista sordida	2	Tricholomataceae	Agaricales	Bas	C	CCBAS761	CCBAS	✓	✓	✓	✓	✓
Leratiomyces laetissimus	1	Strophariaceae	Agaricales	Bas	F	P27	CJB	✓	✓	✓	✓	✓
Leucoagaricus bresadolae	1	Agaricaceae	Agaricales	Bas	C	CCBAS802	CCBAS	✓	✓	✓	✓	✓
Lophiostoma sp.	1	IncertaeSedis	Pleosporales	Asc	C	AK190/05	CFGM	✓	✓	✓	✓	✓
Lycoperdon echinatum	1	Agaricaceae	Agaricales	Bas	F	TZ7	CFGM	–	✓	✓	✓	–
Lycoperdon perlatum	1	Agaricaceae	Agaricales	Bas	C	CCBAS284	CCBAS	✓	✓	✓	✓	✓
Lycoperdon sp.	1	Agaricaceae	Agaricales	Bas	C	CCBAS405	CCBAS	✓	✓	✓	✓	✓
Marasmius alliaceus	3	Omphalotaceae	Agaricales	Bas	C	CCBAS290	CCBAS	✓	✓	✓	✓	✓
Marasmius alliaceus	3	Omphalotaceae	Agaricales	Bas	C	CCBAS290	CCBAS	–	–	✓	✓	✓
Marasmius alliaceus	3	Omphalotaceae	Agaricales	Bas	C	CCBAS623	CCBAS	✓	✓	✓	✓	✓
Marasmius oreades	1	Marasmiaceae	Agaricales	Bas	C	CCBAS353	CCBAS	✓	✓	✓	✓	✓
Megacollybia platyphylla	2	Tricholomataceae	Agaricales	Bas	F	TZ24	CFGM	✓	✓	✓	✓	✓
Megacollybia platyphylla	2	Tricholomataceae	Agaricales	Bas	F	h1	CFGM	✓	✓	✓	✓	✓
Megacollybia rodmanii	1	Tricholomataceae	Agaricales	Bas	F	X46	CJB	✓	✓	✓	✓	✓
Megacollybia sp.	1	Tricholomataceae	Agaricales	Bas	F	X49	CJB	–	✓	–	✓	–
Melanoleuca angelesiana	3	Tricholomataceae	Agaricales	Bas	F	M223	CMT	✓	✓	✓	✓	✓
Melanoleuca angelesiana	3	Tricholomataceae	Agaricales	Bas	F	M240	CMT	✓	✓	✓	✓	✓
Melanoleuca angelesiana	3	Tricholomataceae	Agaricales	Bas	F	M126	CMT	✓	✓	✓	✓	✓
Melanoleuca decembris	3	Tricholomataceae	Agaricales	Bas	F	M108	CMT	✓	✓	✓	✓	✓
Melanoleuca decembris	3	Tricholomataceae	Agaricales	Bas	F	M166	CMT	✓	✓	✓	✓	✓
Melanoleuca decembris	3	Tricholomataceae	Agaricales	Bas	F	M142	CMT	✓	✓	✓	✓	✓
Melanoleuca exscissa	3	Tricholomataceae	Agaricales	Bas	F	M109	CMT	✓	✓	✓	✓	✓
Melanoleuca exscissa	3	Tricholomataceae	Agaricales	Bas	F	M112	CMT	✓	✓	✓	✓	✓
Melanoleuca exscissa	3	Tricholomataceae	Agaricales	Bas	F	M149	CMT	✓	✓	✓	✓	✓
Melanoleuca grammopodia	3	Tricholomataceae	Agaricales	Bas	F	M145	CMT	✓	✓	✓	✓	✓
Melanoleuca grammopodia	3	Tricholomataceae	Agaricales	Bas	F	M85	CMT	✓	✓	✓	✓	✓
Melanoleuca grammopodia	3	Tricholomataceae	Agaricales	Bas	F	M139	CMT	✓	✓	✓	✓	✓

Melanoleuca humilis	5	Tricholomataceae	Agaricales	Bas	F	M82	CMT	✓	✓	✓	✓	✓
Melanoleuca humilis	5	Tricholomataceae	Agaricales	Bas	F	M271	CMT	✓	✓	✓	✓	✓
Melanoleuca humilis	5	Tricholomataceae	Agaricales	Bas	F	M49	CMT	✓	✓	✓	✓	✓
Melanoleuca humilis	5	Tricholomataceae	Agaricales	Bas	F	M226	CMT	✓	✓	✓	✓	✓
Melanoleuca humilis	5	Tricholomataceae	Agaricales	Bas	F	M69	CMT	✓	✓	✓	✓	✓
Melanoleuca sp.1	1	Tricholomataceae	Agaricales	Bas	F	M47	CMT	✓	✓	✓	✓	✓
Melanoleuca sp.2	1	Tricholomataceae	Agaricales	Bas	F	M246	CMT	✓	✓	✓	✓	✓
Melanoleuca sublanipes	3	Tricholomataceae	Agaricales	Bas	F	M44	CMT	✓	✓	✓	✓	✓
Melanoleuca sublanipes	3	Tricholomataceae	Agaricales	Bas	F	M111	CMT	✓	✓	✓	✓	✓
Melanoleuca sublanipes	3	Tricholomataceae	Agaricales	Bas	F	M250	CMT	✓	✓	✓	✓	✓
Mensularia radiata	1	Hymenochaetaceae	Hymenochaetales	Bas	C	H47	CFGM	✓	✓	✓	✓	✓
Meyerozyma guilliermondii	1	Debaryomycetaceae	Saccharomycetales	Asc	C	CCBAS796	CCBAS	✓	✓	✓	✓	✓
Monilinia laxa	1	Sclerotiniaceae	Helotiales	Asc	C	SK278	CFGM	✓	✓	✓	✓	✓
Mortierella sp.	1	Mortierellaceae	Mortierellales	Mor	C	100	CFGM	✓	✓	✓	✓	✓
Mortierella verticillata	1	Mortierellaceae	Mortierellales	Mor	C	DIG3	CFGM	–	✓	✓	✓	–
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS682	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS438	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS732	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS441	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS439	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS431	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS430	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS428	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS436	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS435	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS425	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS425	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS434	CCBAS	✓	✓	✓	✓	✓
Mucidula mucida	14	Physalacriaceae	Agaricales	Bas	C	CCBAS438	CCBAS	✓	✓	✓	✓	✓
Mucor hiemalis	1	Mucoraceae	Mucorales	Muc	C	B4	CFGM	✓	✓	✓	✓	✓

Mucor racemosus	1	Mucoraceae	Mucorales	Muc	C	79	CFGM	✓	✓	✓	✓	✓
Mucor sp.	1	Mucoraceae	Mucorales	Muc	C	CCF3774	CCF	✓	✓	✓	✓	✓
Mycena crocata	1	Tricholomataceae	Agaricales	Bas	C	CCBAS816	CCBAS	✓	✓	✓	✓	✓
Mycena galopus	1	Tricholomataceae	Agaricales	Bas	F	h31	CFGM	–	–	✓	✓	–
Mycena polygramma	2	Tricholomataceae	Agaricales	Bas	C	CCBAS347	CCBAS	✓	✓	✓	✓	✓
Mycena polygramma	2	Tricholomataceae	Agaricales	Bas	C	CCBAS419	CCBAS	✓	✓	✓	✓	✓
Mycena sp.1	1	Tricholomataceae	Agaricales	Bas	F	h71	CFGM	–	✓	✓	✓	–
Mycena sp.2	1	Tricholomataceae	Agaricales	Bas	F	h75	CFGM	✓	✓	✓	✓	✓
Nemania diffusa	1	Xylariaceae	Xylariales	Asc	C	H33	CFGM	✓	✓	✓	✓	✓
Nemania serpens	1	Xylariaceae	Xylariales	Asc	C	H05	CFGM	✓	✓	✓	✓	✓
Obolarina dryophila	1	Xylariaceae	Xylariales	Asc	C	H04	CFGM	✓	✓	✓	✓	✓
Ochrocladosporium sp.	1	IncertaeSedis	Pleosporales	Asc	C	AK9/09	CFGM	✓	✓	✓	✓	✓
Omphalina mutila	1	Tricholomataceae	Agaricales	Bas	C	CCBAS343	CCBAS	✓	✓	✓	✓	✓
Omphalotus japonicus	1	Omphalotaceae	Agaricales	Bas	C	CCBAS388	CCBAS	✓	✓	✓	✓	✓
Ossicaulis lachnopus	2	Lyophyllaceae	Agaricales	Bas	C	PRM899407	PRM	✓	✓	✓	✓	✓
Ossicaulis lachnopus	2	Lyophyllaceae	Agaricales	Bas	C	163/2009	CFGM	✓	✓	✓	✓	✓
Ossicaulis lignatilis	3	Lyophyllaceae	Agaricales	Bas	C	PRM829164	PRM	✓	✓	✓	✓	✓
Ossicaulis lignatilis	3	Lyophyllaceae	Agaricales	Bas	C	148/2011	CFGM	–	✓	✓	–	–
Ossicaulis lignatilis	3	Lyophyllaceae	Agaricales	Bas	C	PRM829198	PRM	✓	✓	✓	✓	✓
Pachylepyrium carbonicola	1	Strophariaceae	Agaricales	Bas	C	CCBAS670	CCBAS	✓	✓	✓	✓	✓
Panaeolus papilionaceus	1	Bolbitiaceae	Agaricales	Bas	F	P10	CJB	✓	✓	✓	✓	✓
Panaeolus sphinctrinus	1	Bolbitiaceae	Agaricales	Bas	F	P16	CJB	–	✓	✓	–	✓
Panellus stipticus	1	Tricholomataceae	Agaricales	Bas	F	TZ21	CFGM	✓	✓	✓	✓	✓
Paraphaeosphaeria michotii	1	Montagnulaceae	Pleosporales	Asc	C	B2C	CFGM	–	✓	✓	✓	–
Paraphaeosphaeria sporulosa	1	Didymosphaeriaceae	Pleosporales	Asc	C	H26	CFGM	✓	✓	✓	✓	✓
Paxillus involutus	1	Paxillaceae	Boletales	Bas	F	TZ27	CFGM	–	–	✓	✓	✓
Penicillium oxalicum	4	Aspergillaceae	Eurotiales	Asc	C	CCF1659	CCF	–	✓	✓	–	✓
Penicillium oxalicum	4	Aspergillaceae	Eurotiales	Asc	C	CCF2315	CCF	–	✓	✓	–	✓
Penicillium oxalicum	4	Aspergillaceae	Eurotiales	Asc	C	CCF1959	CCF	✓	✓	✓	✓	✓
Penicillium oxalicum	4	Aspergillaceae	Eurotiales	Asc	C	AK98_11	CFGM	✓	✓	✓	✓	✓

Penicillium soosianum	1	Aspergillaceae	Eurotiales	Asc	C	MH53	CFGM	✓	✓	✓	✓	✓
Penicillium sp.	1	Aspergillaceae	Eurotiales	Asc	C	VK384	CFGM	–	–	✓	–	–
Petriella sp.	1	Microascaceae	Microascales	Asc	C	AK64/08	CFGM	✓	✓	✓	✓	✓
Peziza arvernensis	1	Pezizaceae	Pezizales	Asc	F	TZ4	CFGM	✓	✓	✓	✓	✓
Phaeolus schweinitzii	1	Hymenochaetaceae	Hymenochaetales	Bas	F	TZ3	CFGM	✓	✓	✓	✓	✓
Phaeomollisia piceae	1	IncertaeSedis	Helotiales	Asc	C	H01	CFGM	–	✓	✓	✓	–
Phanerochaete chrysosporium	1	Phanerochaetaceae	Polyporales	Bas	C	CCBAS570	CCBAS	✓	✓	✓	✓	✓
Phanerochaete sanguinea	1	Phanerochaetaceae	Polyporales	Bas	C	CCBAS845	CCBAS	✓	✓	✓	✓	✓
Phanerochaete sordida	1	Phanerochaetaceae	Polyporales	Bas	C	CCBAS531	CCBAS	✓	✓	✓	✓	✓
Phanerochaete sp.	1	Phanerochaetaceae	Polyporales	Bas	C	CCBAS570	CCBAS	✓	✓	✓	✓	✓
Phellinus ferrugineus	1	Hymenochaetaceae	Hymenochaetales	Bas	C	87sety	CFGM	✓	✓	✓	✓	✓
Phellinus igniarius	3	Hymenochaetaceae	Hymenochaetales	Bas	F	923	CMT	✓	✓	✓	✓	✓
Phellinus igniarius	3	Hymenochaetaceae	Hymenochaetales	Bas	F	984	CMT	✓	✓	✓	✓	✓
Phellinus igniarius	3	Hymenochaetaceae	Hymenochaetales	Bas	C	CCBAS575	CCBAS	✓	✓	✓	✓	✓
Phellinus robustus	1	Hymenochaetaceae	Hymenochaetales	Bas	C	CCBAS574	CCBAS	✓	✓	✓	✓	✓
Phellinus tuberculosus	2	Hymenochaetaceae	Hymenochaetales	Bas	F	985	CMT	–	✓	✓	✓	–
Phellinus tuberculosus	2	Hymenochaetaceae	Hymenochaetales	Bas	F	PH74	CMT	–	✓	✓	✓	–
Phellopilus nigrolimitatus	1	Hymenochaetaceae	Hymenochaetales	Bas	C	CCBAS578	CCBAS	✓	✓	✓	✓	✓
Phialocephala compacta	1	mitosporicHelotiales	Helotiales	Asc	C	H45	CFGM	✓	✓	✓	✓	✓
Phlebia chrysocreas	2	Corticiaceae	Corticiales	Bas	C	CCBAS716	CCBAS	–	✓	–	✓	✓
Phlebia chrysocreas	2	Corticiaceae	Corticiales	Bas	C	CCBAS715	CCBAS	✓	✓	✓	✓	✓
Phlebia sp.	1	Corticiaceae	Corticiales	Bas	C	CCBAS716	CCBAS	✓	✓	✓	✓	✓
Phlebia tremellosa	1	Corticiaceae	Corticiales	Bas	F	h8	CFGM	–	✓	–	✓	✓
Pholiota aurivella	1	Strophariaceae	Agaricales	Bas	C	CCBAS458	CCBAS	✓	✓	✓	✓	✓
Pholiota carbonaria	1	Strophariaceae	Agaricales	Bas	F	PRM888152	PRM	✓	✓	✓	✓	✓
Pholiota highlandensis	1	Strophariaceae	Agaricales	Bas	C	PH895180	CFGM	–	✓	✓	✓	–
Pholiota lubrica	2	Strophariaceae	Agaricales	Bas	C	PRM899117	PRM	✓	✓	✓	✓	✓
Pholiota lubrica	2	Strophariaceae	Agaricales	Bas	C	PRM915546	PRM	✓	✓	✓	✓	✓
Pholiota mixta	1	Strophariaceae	Agaricales	Bas	C	PRM909924	PRM	✓	✓	✓	✓	✓
Pholiota spumosa	1	Strophariaceae	Agaricales	Bas	F	PRM857179	PRM	✓	✓	✓	✓	✓

Pholiota squamosa	2	Strophariaceae	Agaricales	Bas	C	PRM885615	PRM	✓	✓	✓	✓	✓
Pholiota squamosa	2	Strophariaceae	Agaricales	Bas	C	PRM901623	PRM	✓	✓	✓	✓	✓
Pholiota virescentifolia	1	Strophariaceae	Agaricales	Bas	F	PRM897292	PRM	–	✓	✓	✓	–
Phoma herbarum	1	Didymellaceae	Pleosporales	Asc	C	Pstruh330	CFGM	–	✓	✓	✓	–
Phoma macrostoma	1	Didymellaceae	Pleosporales	Asc	C	B5/2	CFGM	✓	✓	✓	✓	✓
Phoma sp.1	1	Didymellaceae	Pleosporales	Asc	C	E38	CFGM	–	✓	✓	✓	–
Phoma sp.2	1	Didymellaceae	Pleosporales	Asc	C	AK195/05	CFGM	✓	✓	✓	✓	✓
Phyllosticta elongata	1	Phyllostictaceae	Botryosphaeriales	Asc	C	CCBAS765	CCBAS	✓	✓	✓	✓	✓
Phyllotopsis nidulans	1	Tricholomataceae	Agaricales	Bas	F	h25	CFGM	–	✓	✓	✓	–
Piptoporus betulinus	1	Coriolaceae	Polyporales	Bas	F	h41	CFGM	✓	✓	✓	✓	✓
Pleosporales sp.1	1	IncertaeSedis	Pleosporales	Asc	C	AK237/05	CFGM	–	✓	✓	✓	–
Pleosporales sp.2	1	IncertaeSedis	Pleosporales	Asc	C	E55	CFGM	✓	✓	✓	✓	✓
Pleurotus calyptratus	1	Pleurotaceae	Agaricales	Bas	C	CCBAS461	CCBAS	✓	✓	✓	✓	✓
Pleurotus cornucopiae	1	Pleurotaceae	Agaricales	Bas	C	CCBAS463	CCBAS	✓	✓	✓	✓	✓
Pleurotus eryngii	4	Pleurotaceae	Agaricales	Bas	C	CCBAS348	CCBAS	✓	✓	✓	✓	✓
Pleurotus eryngii	4	Pleurotaceae	Agaricales	Bas	C	CCBAS347	CCBAS	✓	✓	✓	✓	✓
Pleurotus eryngii	4	Pleurotaceae	Agaricales	Bas	C	CCBAS348	CCBAS	✓	✓	✓	✓	✓
Pleurotus eryngii	4	Pleurotaceae	Agaricales	Bas	C	CCBAS485	CCBAS	✓	✓	✓	✓	✓
Pleurotus ostreatus	1	Pleurotaceae	Agaricales	Bas	C	CCBAS766	CCBAS	✓	✓	✓	✓	✓
Pleurotus pulmonarius	1	Pleurotaceae	Agaricales	Bas	C	CCBAS481	CCBAS	–	✓	✓	✓	–
Pleurotus sp.	1	Pleurotaceae	Agaricales	Bas	C	CCBAS477	CCBAS	✓	✓	✓	✓	✓
Pochonia suchlasporia	1	Clavicipitaceae	Hypocreales	Asc	C	AK70/11	CFGM	✓	✓	✓	✓	✓
Polyporus brumalis	2	Polyporaceae	Polyporales	Bas	C	CCBAS589	CCBAS	✓	✓	✓	✓	✓
Polyporus brumalis	2	Polyporaceae	Polyporales	Bas	C	CCBAS818	CCBAS	✓	✓	✓	✓	✓
Polyporus sp.1	1	Polyporaceae	Polyporales	Bas	C	CCBAS592	CCBAS	✓	✓	✓	✓	✓
Polyporus sp.2	1	Polyporaceae	Polyporales	Bas	C	CCBAS102	CCBAS	✓	✓	✓	✓	✓
Polyporus sp.3	1	Polyporaceae	Polyporales	Bas	C	CCBAS588	CCBAS	✓	✓	✓	✓	✓
Polyporus sp.4	1	Polyporaceae	Polyporales	Bas	C	CCBAS598	CCBAS	✓	✓	✓	✓	✓
Polyporus squamosus	2	Polyporaceae	Polyporales	Bas	C	CCBAS676	CCBAS	✓	✓	✓	✓	✓
Polyporus squamosus	2	Polyporaceae	Polyporales	Bas	C	CCBAS676	CCBAS	✓	✓	✓	✓	✓

Porodaedalea laricis	3	Hymenochaetaceae	Hymenochaetales	Bas	F	PH110	CMT	✓	✓	✓	✓	✓
Porodaedalea laricis	3	Hymenochaetaceae	Hymenochaetales	Bas	F	PH118	CMT	✓	✓	✓	✓	✓
Porodaedalea laricis	3	Hymenochaetaceae	Hymenochaetales	Bas	F	OM10	CMT	✓	✓	✓	✓	✓
Porodaedalea pini	3	Hymenochaetaceae	Hymenochaetales	Bas	F	TM1	CMT	✓	✓	✓	✓	✓
Porodaedalea pini	3	Hymenochaetaceae	Hymenochaetales	Bas	F	U1	CMT	✓	✓	✓	✓	✓
Porodaedalea pini	3	Hymenochaetaceae	Hymenochaetales	Bas	F	1031	CMT	✓	✓	✓	✓	✓
Porodaedalea sp.	1	Hymenochaetaceae	Hymenochaetales	Bas	C	CCBAS735	CCBAS	✓	✓	✓	✓	✓
Psathyrella sp.	1	Psathyrellaceae	Agaricales	Bas	F	h74	CFGM	✓	✓	✓	✓	✓
Pseudobaeospora sp.	1	Tricholomataceae	Agaricales	Bas	F	X17	CJB	✓	✓	✓	✓	✓
Pseudogymnoascus pannorum	1	Pseudeurotiaceae	IncertaeSedis	Asc	C	AK51/11	CFGM	✓	✓	✓	✓	✓
Psilocybe allenii	3	Strophariaceae	Agaricales	Bas	F	P41	CJB	✓	✓	✓	✓	✓
Psilocybe allenii	3	Strophariaceae	Agaricales	Bas	F	P42	CJB	✓	✓	✓	✓	✓
Psilocybe allenii	3	Strophariaceae	Agaricales	Bas	F	P40	CJB	–	✓	✓	–	✓
Psilocybe arcana	1	Strophariaceae	Agaricales	Bas	C	CCBAS714	CCBAS	✓	✓	✓	✓	✓
Psilocybe atrobrunnea	5	Strophariaceae	Agaricales	Bas	F	P26	CJB	✓	✓	✓	✓	✓
Psilocybe atrobrunnea	5	Strophariaceae	Agaricales	Bas	F	P23	CJB	–	✓	✓	✓	–
Psilocybe atrobrunnea	5	Strophariaceae	Agaricales	Bas	F	P32	CJB	–	✓	✓	✓	–
Psilocybe atrobrunnea	5	Strophariaceae	Agaricales	Bas	F	P36	CJB	–	✓	✓	✓	–
Psilocybe atrobrunnea	5	Strophariaceae	Agaricales	Bas	F	P54	CJB	✓	✓	✓	✓	✓
Psilocybe azurescens	3	Strophariaceae	Agaricales	Bas	F	P22	CJB	–	✓	✓	✓	–
Psilocybe azurescens	3	Strophariaceae	Agaricales	Bas	F	P1	CJB	✓	✓	✓	✓	✓
Psilocybe azurescens	3	Strophariaceae	Agaricales	Bas	F	P29	CJB	✓	✓	✓	✓	✓
Psilocybe bohemica	2	Strophariaceae	Agaricales	Bas	F	P12	CJB	✓	✓	✓	✓	✓
Psilocybe bohemica	2	Strophariaceae	Agaricales	Bas	F	P13	CJB	✓	✓	✓	✓	✓
Psilocybe caerulipes	2	Strophariaceae	Agaricales	Bas	F	P51	CJB	✓	✓	✓	✓	✓
Psilocybe caerulipes	2	Strophariaceae	Agaricales	Bas	F	P53	CJB	✓	✓	✓	✓	✓
Psilocybe cubensis	1	Strophariaceae	Agaricales	Bas	F	P30	CJB	✓	✓	✓	✓	✓
Psilocybe cyanescens	3	Strophariaceae	Agaricales	Bas	F	P21	CJB	✓	✓	✓	✓	✓
Psilocybe cyanescens	3	Strophariaceae	Agaricales	Bas	C	CCBAS820	CCBAS	✓	✓	✓	✓	✓
Psilocybe cyanescens	3	Strophariaceae	Agaricales	Bas	F	P39	CJB	✓	✓	✓	✓	✓

Psilocybe luteonitens	1	Strophariaceae	Agaricales	Bas	F	P50	CJB	✓	✓	✓	✓	✓
Psilocybe medullosa	3	Strophariaceae	Agaricales	Bas	F	P17	CJB	✓	✓	✓	✓	✓
Psilocybe medullosa	3	Strophariaceae	Agaricales	Bas	F	P37	CJB	✓	✓	✓	✓	✓
Psilocybe medullosa	3	Strophariaceae	Agaricales	Bas	F	P11	CJB	✓	✓	✓	✓	✓
Psilocybe moravica	1	Strophariaceae	Agaricales	Bas	F	P6	CJB	–	✓	✓	✓	–
Psilocybe semilanceata	2	Strophariaceae	Agaricales	Bas	F	P31	CJB	✓	✓	✓	✓	✓
Psilocybe semilanceata	2	Strophariaceae	Agaricales	Bas	C	CCBAS494	CCBAS	✓	✓	✓	✓	✓
Psilocybe serbica	1	Strophariaceae	Agaricales	Bas	F	P15	CJB	✓	✓	✓	✓	✓
Psilocybe silvatica	1	Strophariaceae	Agaricales	Bas	F	P24	CJB	✓	✓	✓	✓	✓
Psilocybe sp.	1	Strophariaceae	Agaricales	Bas	F	P8	CJB	–	✓	✓	✓	–
Psilocybe squamosa	1	Strophariaceae	Agaricales	Bas	F	P38	CJB	✓	✓	✓	✓	✓
Psilocybe subaeruginosa	4	Strophariaceae	Agaricales	Bas	F	P43	CJB	✓	✓	✓	✓	✓
Psilocybe subaeruginosa	4	Strophariaceae	Agaricales	Bas	C	CCBAS488	CCBAS	✓	✓	✓	✓	✓
Psilocybe subaeruginosa	4	Strophariaceae	Agaricales	Bas	F	P44	CJB	✓	✓	✓	✓	✓
Psilocybe subaeruginosa	4	Strophariaceae	Agaricales	Bas	F	P46	CJB	✓	✓	✓	✓	✓
Psilocybe weraroa	2	Strophariaceae	Agaricales	Bas	F	P34	CJB	–	✓	✓	✓	–
Psilocybe weraroa	2	Strophariaceae	Agaricales	Bas	F	P35	CJB	–	✓	✓	✓	–
Pyrenochaeta sp.	1	Cucurbitariaceae	Pleosporales	Asc	C	E42	CFGM	✓	✓	✓	✓	✓
Pyronemataceae sp.	1	Pyronemataceae	Pezizales	Asc	C	H19	CFGM	✓	✓	✓	✓	✓
Raffaelea scolytodis	1	Ophiostomataceae	Ophiostomatales	Asc	C	CCF3572	CCF	✓	✓	✓	✓	✓
Ramariopsis helvola	1	Clavariaceae	Agaricales	Bas	F	h12	CFGM	✓	✓	✓	✓	✓
Rhinocladiella atrovirens	1	Herpotrichiellaceae	Chaetothyriales	Asc	C	409c	CFGM	–	–	–	–	–
Rhizopus sp.	1	Rhizopodaceae	Mucorales	Muc	C	289	CFGM	–	–	–	✓	✓
Rhodocollybia butyracea	2	Omphalotaceae	Agaricales	Bas	C	CCBAS279	CCBAS	✓	✓	✓	✓	✓
Rhodocollybia butyracea	2	Omphalotaceae	Agaricales	Bas	F	h43	CFGM	–	–	✓	✓	✓
Rhodocollybia maculata	2	Omphalotaceae	Agaricales	Bas	C	CCBAS349	CCBAS	✓	✓	✓	✓	✓
Rhodocollybia maculata	2	Omphalotaceae	Agaricales	Bas	F	TZ25	CFGM	✓	✓	✓	✓	✓
Russula acrifolia	1	Russulaceae	Russulales	Bas	F	R119	CJB	✓	✓	✓	✓	✓
Russula adusta	3	Russulaceae	Russulales	Bas	F	R99	CJB	✓	✓	✓	✓	✓
Russula adusta	3	Russulaceae	Russulales	Bas	F	R86	CJB	✓	✓	✓	✓	✓

Russula adusta	3	Russulaceae	Russulales	Bas	F	R102	CJB	✓	✓	✓	✓	✓
Russula albonigra	2	Russulaceae	Russulales	Bas	F	R101	CJB	✓	✓	✓	✓	✓
Russula albonigra	2	Russulaceae	Russulales	Bas	F	R114	CJB	✓	✓	✓	✓	✓
Russula alnetorum	3	Russulaceae	Russulales	Bas	F	R82	CJB	✓	✓	✓	✓	✓
Russula alnetorum	3	Russulaceae	Russulales	Bas	F	R7	CJB	✓	✓	✓	✓	✓
Russula alnetorum	3	Russulaceae	Russulales	Bas	F	R9	CJB	✓	✓	✓	✓	✓
Russula alnicrispae	2	Russulaceae	Russulales	Bas	F	R80	CJB	✓	✓	✓	✓	✓
Russula alnicrispae	2	Russulaceae	Russulales	Bas	F	R83	CJB	✓	✓	✓	✓	✓
Russula alutacea	1	Russulaceae	Russulales	Bas	F	R17	CJB	–	✓	–	–	–
Russula amethystina	2	Russulaceae	Russulales	Bas	F	R67	CJB	✓	✓	✓	✓	✓
Russula amethystina	2	Russulaceae	Russulales	Bas	F	R58	CJB	✓	✓	✓	✓	✓
Russula amoenolens	2	Russulaceae	Russulales	Bas	F	R105	CJB	✓	✓	✓	✓	✓
Russula amoenolens	2	Russulaceae	Russulales	Bas	F	R84	CJB	✓	✓	✓	✓	✓
Russula anthracina	1	Russulaceae	Russulales	Bas	F	R57	CJB	✓	✓	✓	✓	✓
Russula atropurpurea	2	Russulaceae	Russulales	Bas	F	R122	CJB	✓	✓	✓	✓	✓
Russula atropurpurea	2	Russulaceae	Russulales	Bas	F	R123	CJB	✓	✓	✓	✓	✓
Russula brevipes	1	Russulaceae	Russulales	Bas	F	R87	CJB	✓	✓	✓	✓	✓
Russula brunneoviolacea	1	Russulaceae	Russulales	Bas	F	R52	CJB	✓	✓	✓	✓	✓
Russula cessans	2	Russulaceae	Russulales	Bas	F	R92	CJB	✓	✓	✓	✓	✓
Russula cessans	2	Russulaceae	Russulales	Bas	F	R59	CJB	✓	✓	✓	✓	✓
Russula chloroides	1	Russulaceae	Russulales	Bas	F	R88	CJB	✓	✓	✓	✓	✓
Russula densifolia	2	Russulaceae	Russulales	Bas	F	R69	CJB	✓	✓	✓	✓	✓
Russula densifolia	2	Russulaceae	Russulales	Bas	F	R77	CJB	✓	✓	✓	✓	✓
Russula foetens	6	Russulaceae	Russulales	Bas	F	R31	CJB	✓	✓	✓	✓	✓
Russula foetens	6	Russulaceae	Russulales	Bas	F	R27	CJB	✓	✓	✓	✓	✓
Russula foetens	6	Russulaceae	Russulales	Bas	F	R24	CJB	✓	✓	✓	✓	✓
Russula foetens	6	Russulaceae	Russulales	Bas	F	R26	CJB	✓	✓	✓	✓	✓
Russula foetens	6	Russulaceae	Russulales	Bas	F	R32	CJB	✓	✓	✓	✓	✓
Russula foetens	6	Russulaceae	Russulales	Bas	F	R37	CJB	–	–	✓	✓	✓
Russula foetentula	1	Russulaceae	Russulales	Bas	F	R79	CJB	✓	✓	✓	✓	✓

Russula fragilis	1	Russulaceae	Russulales	Bas	F	R113	CJB	✓	✓	✓	✓	✓
Russula fragrantissima	2	Russulaceae	Russulales	Bas	F	R100	CJB	✓	✓	✓	✓	✓
Russula fragrantissima	2	Russulaceae	Russulales	Bas	F	R103	CJB	✓	✓	✓	✓	✓
Russula globispora	1	Russulaceae	Russulales	Bas	F	R15	CJB	–	✓	✓	✓	–
Russula helodes	2	Russulaceae	Russulales	Bas	F	R97	CJB	✓	✓	✓	✓	✓
Russula helodes	2	Russulaceae	Russulales	Bas	F	R98	CJB	✓	✓	✓	✓	✓
Russula hortensis	1	Russulaceae	Russulales	Bas	F	R72	CJB	✓	✓	✓	✓	✓
Russula ilicis	1	Russulaceae	Russulales	Bas	F	R115	CJB	✓	✓	✓	✓	✓
Russula illota	4	Russulaceae	Russulales	Bas	F	R29	CJB	✓	✓	✓	✓	✓
Russula illota	4	Russulaceae	Russulales	Bas	F	R65	CJB	✓	✓	✓	✓	✓
Russula illota	4	Russulaceae	Russulales	Bas	F	R106	CJB	✓	✓	✓	✓	✓
Russula illota	4	Russulaceae	Russulales	Bas	F	R107	CJB	✓	✓	✓	✓	✓
Russula insignis	1	Russulaceae	Russulales	Bas	F	R76	CJB	✓	✓	✓	✓	✓
Russula integra	3	Russulaceae	Russulales	Bas	F	R71	CJB	✓	✓	✓	✓	✓
Russula integra	3	Russulaceae	Russulales	Bas	F	R48	CJB	✓	✓	✓	✓	✓
Russula integra	3	Russulaceae	Russulales	Bas	F	R89	CJB	✓	✓	✓	✓	✓
Russula laeta	1	Russulaceae	Russulales	Bas	F	R70	CJB	–	–	–	–	✓
Russula maculata	1	Russulaceae	Russulales	Bas	F	R66	CJB	–	✓	–	✓	✓
Russula mutabilis	1	Russulaceae	Russulales	Bas	F	R121	CJB	✓	✓	✓	✓	✓
Russula nigricans	1	Russulaceae	Russulales	Bas	F	R95	CJB	✓	✓	✓	✓	✓
Russula nitida	1	Russulaceae	Russulales	Bas	F	R91	CJB	✓	✓	✓	✓	✓
Russula ochroleuca	2	Russulaceae	Russulales	Bas	F	R90	CJB	✓	✓	✓	✓	✓
Russula ochroleuca	2	Russulaceae	Russulales	Bas	F	R125	CJB	✓	✓	✓	✓	✓
Russula pascua	1	Russulaceae	Russulales	Bas	F	h51	CFGM	✓	✓	✓	✓	✓
Russula pectinatoides	5	Russulaceae	Russulales	Bas	F	R116	CJB	✓	✓	✓	✓	✓
Russula pectinatoides	5	Russulaceae	Russulales	Bas	F	R93	CJB	✓	✓	✓	✓	✓
Russula pectinatoides	5	Russulaceae	Russulales	Bas	F	R109	CJB	✓	✓	✓	✓	✓
Russula pectinatoides	5	Russulaceae	Russulales	Bas	F	R110	CJB	✓	✓	✓	✓	✓
Russula pectinatoides	5	Russulaceae	Russulales	Bas	F	R111	CJB	✓	✓	✓	✓	✓
Russula pumila	6	Russulaceae	Russulales	Bas	F	R6	CJB	✓	✓	✓	✓	✓

Russula pumila	6	Russulaceae	Russulales	Bas	F	R2	CJB	✓	✓	✓	✓	✓
Russula pumila	6	Russulaceae	Russulales	Bas	F	R3	CJB	✓	✓	✓	✓	✓
Russula pumila	6	Russulaceae	Russulales	Bas	F	R1	CJB	✓	✓	✓	✓	✓
Russula pumila	6	Russulaceae	Russulales	Bas	F	R5	CJB	✓	✓	✓	✓	✓
Russula pumila	6	Russulaceae	Russulales	Bas	F	R4	CJB	✓	✓	✓	✓	✓
Russula putida	1	Russulaceae	Russulales	Bas	F	R35	CJB	✓	✓	✓	✓	✓
Russula rhodopoda	3	Russulaceae	Russulales	Bas	F	R94	CJB	✓	✓	✓	✓	✓
Russula rhodopoda	3	Russulaceae	Russulales	Bas	F	R13	CJB	✓	✓	✓	✓	✓
Russula rhodopoda	3	Russulaceae	Russulales	Bas	F	R21	CJB	–	–	–	✓	✓
Russula roberti	1	Russulaceae	Russulales	Bas	F	R46	CJB	✓	✓	✓	✓	✓
Russula rubra	1	Russulaceae	Russulales	Bas	F	R120	CJB	–	✓	–	✓	✓
Russula rubroalba	1	Russulaceae	Russulales	Bas	F	R117	CJB	✓	✓	✓	✓	✓
Russula seperina	2	Russulaceae	Russulales	Bas	F	R14	CJB	✓	✓	✓	✓	✓
Russula seperina	2	Russulaceae	Russulales	Bas	F	R19	CJB	✓	✓	✓	✓	✓
Russula sororia	2	Russulaceae	Russulales	Bas	F	R104	CJB	✓	✓	✓	✓	✓
Russula sororia	2	Russulaceae	Russulales	Bas	F	R85	CJB	✓	✓	✓	✓	✓
Russula sp.1	1	Russulaceae	Russulales	Bas	F	R108	CJB	✓	✓	✓	✓	✓
Russula sp.2	1	Russulaceae	Russulales	Bas	F	R112	CJB	✓	✓	✓	✓	✓
Russula sphagnophila	2	Russulaceae	Russulales	Bas	F	R96	CJB	✓	✓	✓	✓	✓
Russula sphagnophila	2	Russulaceae	Russulales	Bas	F	R47	CJB	✓	✓	✓	✓	✓
Russula subarctica	1	Russulaceae	Russulales	Bas	F	R81	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R30	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R23	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R62	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R42	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R28	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R22	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R25	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R43	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R53	CJB	✓	✓	✓	✓	✓

Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R54	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R60	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R78	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R118	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R38	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R39	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R40	CJB	✓	✓	✓	✓	✓
Russula subfoetens	17	Russulaceae	Russulales	Bas	F	R61	CJB	✓	✓	✓	✓	✓
Russula turci	4	Russulaceae	Russulales	Bas	F	R50	CJB	✓	✓	✓	✓	✓
Russula turci	4	Russulaceae	Russulales	Bas	F	R63	CJB	✓	✓	✓	✓	✓
Russula turci	4	Russulaceae	Russulales	Bas	F	R44	CJB	✓	✓	✓	✓	✓
Russula turci	4	Russulaceae	Russulales	Bas	F	R49	CJB	✓	✓	✓	✓	✓
Russula undulata	1	Russulaceae	Russulales	Bas	F	R51	CJB	✓	✓	✓	✓	✓
Russula versatilis	1	Russulaceae	Russulales	Bas	F	R55	CJB	✓	✓	✓	✓	✓
Russula vinosobrunnea	1	Russulaceae	Russulales	Bas	F	R16	CJB	–	–	✓	–	✓
Russula virescens	1	Russulaceae	Russulales	Bas	F	TZ15	CFGM	–	–	✓	✓	✓
Russula viscida	1	Russulaceae	Russulales	Bas	F	R124	CJB	✓	✓	✓	✓	✓
Schizophyllum commune	6	Schizophyllaceae	Agaricales	Bas	C	CCBAS600	CCBAS	✓	✓	✓	✓	✓
Schizophyllum commune	6	Schizophyllaceae	Agaricales	Bas	C	CCBAS752	CCBAS	✓	✓	✓	✓	✓
Schizophyllum commune	6	Schizophyllaceae	Agaricales	Bas	C	CCBAS602	CCBAS	✓	✓	✓	✓	✓
Schizophyllum commune	6	Schizophyllaceae	Agaricales	Bas	C	CCF3578	CCF	✓	✓	✓	✓	✓
Schizophyllum commune	6	Schizophyllaceae	Agaricales	Bas	C	H31	CFGM	✓	✓	✓	✓	✓
Schizophyllum commune	6	Schizophyllaceae	Agaricales	Bas	C	CCBAS600	CCBAS	✓	✓	✓	✓	✓
Schizophyllum radiatum	2	Schizophyllaceae	Agaricales	Bas	C	CCBAS601	CCBAS	✓	✓	✓	✓	✓
Schizophyllum radiatum	2	Schizophyllaceae	Agaricales	Bas	C	CCBAS852	CCBAS	✓	✓	✓	✓	✓
Schizophyllum sp.1	1	Schizophyllaceae	Agaricales	Bas	C	CCBAS603	CCBAS	✓	✓	✓	✓	✓
Schizophyllum sp.2	1	Schizophyllaceae	Agaricales	Bas	C	CCBAS752	CCBAS	✓	✓	✓	✓	✓
Scleroderma citrinum	1	Sclerodermataceae	Boletales	Bas	F	TZ20	CFGM	✓	✓	✓	✓	✓
Scopulariopsis chartarum	1	Microascaceae	Microascales	Asc	C	407A	CFGM	–	✓	✓	✓	–
Serpula himantioides	1	Serpulaceae	Boletales	Bas	C	CCBAS110	CCBAS	✓	✓	✓	✓	✓

Simplicillium lamellicola	1	Cordycipitaceae	Hypocreales	Asc	C	CCBAS401	CCBAS	✓	✓	✓	✓	✓
Sistotrema brinkmannii	1	Corticiaceae	Corticiales	Bas	C	H44	CFGM	✓	✓	✓	✓	✓
Sparassis crispa	1	Sparassidaceae	Polyporales	Bas	C	CCBAS658	CCBAS	–	✓	✓	✓	–
Spongipellis spumeus	1	Polyporaceae	Polyporales	Bas	F	S4	CMT	–	✓	✓	✓	–
Staphylotrichum coccosporum	1	IncertaeSedis	IncertaeSedis	Asc	C	CCF1053	CCF	✓	✓	✓	✓	✓
Staphylotrichum sp.	1	IncertaeSedis	IncertaeSedis	Asc	C	STap2	CFGM	✓	✓	✓	✓	✓
Stereum hirsutum	1	Stereaceae	Russulales	Bas	C	CCBAS525	CCBAS	✓	✓	✓	✓	✓
Strobilurus esculentus	1	Physalacriaceae	Agaricales	Bas	F	h54	CFGM	✓	✓	✓	✓	✓
Stropharia aeruginosa	1	Strophariaceae	Agaricales	Bas	F	h15	CFGM	–	✓	✓	–	✓
Stropharia hemiglobata	1	Strophariaceae	Agaricales	Bas	F	h9	CFGM	✓	✓	✓	✓	✓
Suillus grevillei	1	Suillaceae	Boletales	Bas	F	TZ19	CFGM	✓	✓	✓	✓	✓
Suillus variegatus	1	Suillaceae	Boletales	Bas	F	h40	CFGM	✓	✓	✓	✓	✓
Sydowia polyspora	1	Dothioraceae	Dothideales	Asc	C	B3C	CFGM	✓	✓	✓	✓	✓
Sydowia sp.	1	Dothioraceae	Dothideales	Asc	C	H23	CFGM	✓	✓	✓	✓	✓
Tapinella atrotomentosa	1	Tapinellaceae	Boletales	Bas	F	TZ10	CFGM	✓	✓	✓	✓	✓
Thanatephorus cucumeris	1	Ceratobasidiaceae	Cantharellales	Bas	C	CCBAS230	CCBAS	–	✓	✓	✓	–
Torrubiella sp.1	1	Clavicipitaceae	Hypocreales	Asc	C	CCBAS769	CCBAS	✓	✓	✓	✓	✓
Torrubiella sp.2	1	Clavicipitaceae	Hypocreales	Asc	C	CCBAS281	CCBAS	✓	✓	✓	✓	✓
Trametes hirsuta	1	Coriolaceae	Polyporales	Bas	C	CCBAS528	CCBAS	✓	✓	✓	✓	✓
Trametes sanguinea	2	Coriolaceae	Polyporales	Bas	C	CCBAS596	CCBAS	–	✓	✓	✓	–
Trametes sanguinea	2	Coriolaceae	Polyporales	Bas	C	CCBAS595	CCBAS	✓	✓	✓	✓	✓
Trametes versicolor	1	Coriolaceae	Polyporales	Bas	C	CCBAS200	CCBAS	✓	✓	✓	✓	✓
Trichoderma harzianum	1	Hypocreaceae	Hypocreales	Asc	C	CCM8008	CFGM	✓	✓	✓	✓	✓
Tricholoma mongolicum	1	Tricholomataceae	Agaricales	Bas	C	CCBAS790	CCBAS	✓	✓	✓	✓	✓
Tricholoma sejunctum	1	Tricholomataceae	Agaricales	Bas	C	CCBAS382	CCBAS	✓	✓	✓	✓	✓
Tricholoma sulphureum	1	Tricholomataceae	Agaricales	Bas	F	TZ31	CFGM	–	✓	✓	✓	–
Tricholomataceae sp.	1	Tricholomataceae	Agaricales	Bas	C	AK176/08	CFGM	✓	✓	✓	✓	✓
Tricholomopsis decora	1	Tricholomataceae	Agaricales	Bas	C	PRM899160	PRM	✓	✓	✓	✓	✓
Tricholomopsis flammula	3	Tricholomataceae	Agaricales	Bas	C	140/2009	CFGM	✓	✓	✓	✓	✓
Tricholomopsis flammula	3	Tricholomataceae	Agaricales	Bas	C	162/2009	CFGM	✓	✓	✓	✓	✓

Tricholomopsis flammula	3	Tricholomataceae	Agaricales	Bas	C	172/2009	CFGM	✓	✓	✓	✓	✓
Tricholomopsis osiliensis	2	Tricholomataceae	Agaricales	Bas	C	166/2009	CFGM	–	–	✓	✓	✓
Tricholomopsis osiliensis	2	Tricholomataceae	Agaricales	Bas	C	TO	CFGM	–	–	✓	✓	–
Tricholomopsis rutilans	1	Tricholomataceae	Agaricales	Bas	C	91/2009	CFGM	✓	✓	✓	✓	✓
Tricholomopsis sulphurea	1	Tricholomataceae	Agaricales	Bas	C	T8	CFGM	✓	✓	✓	✓	✓
Trichomonascus farinosus	1	Trichomonascaceae	Saccharomycetales	Asc	C	SF	CFGM	✓	✓	✓	✓	✓
Trichophyton terrestre	1	Arthrodermataceae	Onygenales	Asc	C	AK44/09	CFGM	✓	✓	✓	✓	✓
Trichurus terrophilus	1	Microascaceae	Microascales	Asc	C	CCF3726	CCF	–	✓	✓	✓	–
Umbelopsis isabellina	2	Umbelopsidaceae	Mucorales	Muc	C	CCF2411	CCF	✓	✓	✓	✓	✓
Umbelopsis isabellina	2	Umbelopsidaceae	Mucorales	Muc	C	CCF2412	CCF	✓	✓	✓	✓	✓
Umbelopsis ramanniana	2	Umbelopsidaceae	Mucorales	Muc	C	CCF2805	CCF	–	✓	✓	✓	–
Umbelopsis ramanniana	2	Umbelopsidaceae	Mucorales	Muc	C	277	CFGM	✓	✓	✓	✓	✓
Umbelopsis versiformis	1	Umbelopsidaceae	Mucorales	Muc	C	CCF3263	CCF	–	✓	✓	✓	–
Vascellum pratense	1	Agaricaceae	Agaricales	Bas	F	h62	CFGM	–	✓	✓	✓	–
Verticillium sp.	1	mitosporicHypocreales	Hypocreales	Asc	C	CCBAS448	CCBAS	✓	✓	✓	✓	✓
Weraroa virescens	1	Strophariaceae	Agaricales	Bas	F	P47	CJB	–	✓	✓	–	–
Wolfiporia dilatohypha	2	Polyporaceae	Polyporales	Bas	F	L26	CMT	✓	✓	✓	✓	✓
Wolfiporia dilatohypha	2	Polyporaceae	Polyporales	Bas	F	L27	CMT	✓	✓	✓	✓	✓
Wolfiporia sp.	1	Polyporaceae	Polyporales	Bas	F	L21B	CMT	–	✓	✓	✓	–
Xerocomus badius	1	Boletaceae	Boletales	Bas	F	TZ18	CFGM	✓	✓	✓	✓	✓
Xerocomus pruinatus	1	Boletaceae	Boletales	Bas	F	h3	CFGM	✓	✓	✓	✓	✓
Xeromphalina campanella	1	Tricholomataceae	Agaricales	Bas	F	TZ14	CFGM	✓	✓	✓	✓	✓
Xylariaceae sp.1	1	Xylariaceae	Xylariales	Asc	C	AK152/08	CFGM	✓	✓	✓	✓	✓
Xylariaceae sp.2	1	Xylariaceae	Xylariales	Asc	C	H39	CFGM	✓	✓	✓	✓	✓
Zygorhynchus sp.	1	Mucoraceae	Mucorales	Muc	C	117	CFGM	–	✓	✓	✓	–

9.2 Submitted Manuscript

Complete manuscript with Supplementary files is provided on attached CD.

The *rpb2* gene represents a viable alternative molecular marker for the analysis of environmental fungal communities

Tomáš Větrovský, Miroslav Kolařík, Lucia Žifčáková, **Tomáš Zelenka**, Petr Baldrian

Submitted Manuscript

Journal:	*Molecular Ecology Resources*
Manuscript ID:	MER-15-0170
Manuscript Type:	Resource Article
Date Submitted by the Author:	11-May-2015
Complete List of Authors:	Větrovský, Tomáš Kolařík, Miroslav Žifčáková, Lucia Zelenka, Tomáš Baldrian, Petr; Institute of Microbiology of the ASCR, Laboratory of Environmental Microbiology
Keywords:	Fungi, DNA Barcoding, Community Ecology, Metagenomics

The *rpb2* gene represents a viable alternative molecular marker for the analysis of environmental fungal communities

Tomáš Větrovský[1], Miroslav Kolařík[1], Lucia Žifčáková, **Tomáš Zelenka**, Petr Baldrian

Institute of Microbiology of the ASCR, v.v.i., Vídeňská 1083, 14220 Praha 4, Czech Republic

[1]These authors contributed equally to this work

Keywords: fungi / community composition / diversity / internal transcribed spacer / phylogeny / RNA polymerase II

Corresponding author: Petr Baldrian, Institute of Microbiology of the ASCR, v.v.i., Vídeňská 1083, 14220 Praha 4, Czech Republic; phone: +420 723 770 570, fax: +420 241 062 384, e-mail: baldrian@biomed.cas.cz

Running title: Fungal community analysis using *rpb2* and ITS

Abstract

Although the commonly used internal transcribed spacer region of rDNA (ITS) is well suited for taxonomic identification of fungi, the information on the relative abundance of taxa and diversity is negatively affected by the multi-copy nature of rDNA and the existence of ITS paralogues. Moreover, due to high variability, ITS sequences cannot be used for phylogenetic analyses of unrelated taxa. The part of single-copy gene encoding the second largest subunit of RNA polymerase II (rpb2) was thus compared with first spacer of ITS as an alternative marker for the analysis of fungal communities in spruce forest topsoil and their applicability was tested on a comprehensive mock community. In soil, rpb2 exhibited broad taxonomic coverage of the entire fungal tree of life including basal fungal lineages. The gene exhibited sufficient variation for the use in phylogenetic analyses and taxonomic assignments, although it amplifies also orthologues. The fungal taxon spectra obtained with rbp2 region and ITS1 corresponded, but sequence abundances differed widely, especially in the basal lineages. The proportions of OTU counts and read counts of major fungal groups were close to the reality when rpb2 was used as a molecular marker while they were strongly biased towards the Basidiomycota when using the ITS primers ITS1/ITS4. Although the taxonomic placement of rbp2 sequences is currently more difficult than of the ITS sequences, its discriminative power, quantitative representation of community composition and suitability for phylogenetic analyses represent significant advantages.

Introduction

The advent of next generation sequencing (NGS) of environmental samples has broadened the exploration of fungal diversity and ecology. One of the most widely used applications of NGS is metabarcoding, which uses a selected DNA sequence to catalogue sequence reads into operational taxonomic units (OTUs) and helps to identify the closest known species. Despite its technical feasibility, this approach to diversity exploration is still limited by several factors, including the method of DNA isolation, selection of marker and PCR primers, and analysis of the data (Bellemain *et al.* 2010; Engelbrektson *et al.* 2010; Gihring *et al.* 2012; Huse *et al.* 2010; Kunin *et al.* 2010; Quince *et al.* 2008; Schloss 2010; Youssef *et al.* 2009). Since the first mycological study applying NGS (Öpik *et al.* 2009), parts of the rDNA cluster have been exclusively used as a molecular marker (Větrovský & Baldrian 2013). The ITS rDNA region, accessible with universal primers, can sufficiently distinguish between most fungal species and is also the most abundant fungal marker in public databases (Nilsson *et al.* 2008; Schoch *et al.* 2012). The rDNA cluster is a multi-copy marker present in fungal genomes in 1-200 copies (Baldrian *et al.* 2013; Debode *et al.* 2009; Herrera *et al.* 2009; Raidl *et al.* 2005); although the multi-copy nature of this sequence facilitates obtaining ITS amplicons from low-quality DNA, it also seriously limits its value for the quantification of the relative abundance of fungal taxa. The multi-copy nature also results in intraspecies and intragenomic variability. Comprehensive studies mapping the distribution of intragenomic ITS rDNA variability among fungi are lacking, though is extent is generally considered to be low (Lindahl *et al.* 2013; Lindner *et al.* 2013) with the exception of the *Glomeromycota* (Krueger *et al.* 2012, Stockinger *et al.* 2010). Most of the studies on the *Ascomycota* and the *Basidiomycota* (Connell *et al.* 2010; Lindner *et al.* 2013; Simon & Weiss 2008; Smith *et al.* 2007; Wang & Yao 2005) have reported a within-genome variability lower than 3%, which corresponds to the typical intraspecies variation in this group. In addition to these shallow levels of paralogy, however, deep paralogues reaching 3.6% (Simon & Weiss 2008) or even 10% dissimilarity (Lindner & Banik 2011; Lindner *et al.* 2013) have also been detected. Even more variability could be hidden in potentially unrecognised rDNA pseudogenes (reviewed in Glass *et al.* 2013).

ITS rDNA is a quantitatively dominant marker in public databases, followed by β-tubulin (*tub2*), translation elongation factor 1-α (*tef1α*), and the second largest subunit of ribosomal polymerase II (*rpb2*) (Feau *et al.* 2011), and the last two also represent the

markers included in the Assembling the Fungal Tree of Life (AFTOL) project (Lutzoni *et al.* 2004, James *et al.* 2006). Of them, *tef1α* is not universaly present in fungi and together with *tub2* is known to have paralogous copies in certain fungal genomes (Keeling & Inagaki 2004, James *et al.* 2006, Hubka & Kolařík 2012). Thus, the *rpb2* gene is a suitable alternative marker characterised in 6378 from 30780 species deposited in NCBI Genbank (Table S1, Supporting Information). ITS rDNA was proposed as the universal fungal barcode (Schoch *et al.* 2012) for its interspecific variability and the availability of conserved primer sites. However, its intragenomic variability could represent a potential limitation for the study of fungi from environmental DNA (Kiss 2012). Cloning or massively parallel sequencing, unlike Sanger sequencing from genomic DNA, yields sequences derived from single alleles, and fungal diversity may be substantially overestimated if these sequences are sufficiently different within a fungal individual or taxon (e.g., when deep paralogues or pseudogenes are recovered) (Lindner & Banik 2011, Lucking *et al.* 2013). The variable length of the ITS region represents another important problem because there is a strong PCR bias against species with longer amplicons, an issue that was found to largely affect the results of community studies (Ihrmark *et al.* 2012). Lastly, the low conservancy of the ITS region precludes its use in phylogenetic studies on higher taxonomic ranks. The protein-encoding genes are suitable for this, since the translated amino acid sequences can be utilised for high-quality alignments of unrelated fungi. This could be exceptionally valuable for the proper placement of unknown higher taxonomic level lineages, which are often encountered in environmental samples (Baldrian *et al.* 2012; Glass *et al.* 2013).

For the above reasons, the use of single-copy protein-encoding genes as alternative markers could be a solution to the multiple problems associated with rDNA-based markers. Although it is recognised that protein-encoding markers, including the *rpb2* gene, have a better species-resolving power than rDNA markers, the absence of universal primers was believed to limit their use as potential universal barcodes (Schoch *et al.* 2012). In this study, we describe the use of the *rpb2* gene as an alternative marker for fungal community analyses in comparison to the ITS region for investigating fungi inhabiting a coniferous forest floor. Forest litter is dominated by mycorrhizal and saprotrophic basidiomycetes (Baldrian *et al.* 2012), both of which are groups with potentially high intragenomic rDNA variation (Kåuserud & Schumacher 2003; Lindner & Banik 2011; Smith *et al.* 2007). In addition, the *in vitro* assembled mock-community consisting of the DNA of 130 species was analysed for comparison. Our aim was to

explore the potential of the *rpb2* gene as an alternative molecular barcode for the study of fungal diversity and ecology in environmental DNA samples.

Materials and Methods

Study site and sample collection

The study area was located in the highest altitudes (1170–1200 m) of the Bohemian Forest mountain range (Central Europe) and was covered by an unmanaged spruce (*Picea abies*) forest (49°02.64 N, 13°37.01 E), an area previously used for the study of active and total microbial community composition (Baldrian *et al.* 2012). To reduce seasonal effects on the fungal community composition, the sampling was performed in September 2010 (autumn) and March 2011 (spring); the same three sites, located 250 m from each other, were sampled on both occasions. Six topsoil samples were collected around the circumference of a 4-m-diameter circle for each of sampling sites (24 samples in total). The litter horizon (L) and organic soil horizon (S) materials were separately pooled for each site. After the removal of roots, the L material was cut into 0.5-cm pieces and mixed; the S material was passed through a 5-mm sterile mesh and mixed. Samples to be used for DNA extraction were immediately frozen in liquid nitrogen and stored on dry ice.

DNA was extracted in triplicate from 0.300-g portions of each sample using the SV method (Sagova-Mareckova *et al.* 2008) and cleaned using the Geneclean Turbo Kit (MP Bioiomedicals, Solon, OH, USA). A solution (1 M HEPES / 1 M $CaCl_2$, pH 7) was added prior to the cleaning procedure, the sample was left standing for 5 min, and the manufacturer's instructions were then followed.

Tag-encoded amplicon pyrosequencing of soil fungal community

PCR reactions were performed independently for each extracted DNA sample. The volume of each PCR sample was 50 µl. PCR primers ITS1 (TCCGTAGGTGAACCTGCGG) / ITS4 (TCCTCCGCTTATTGATATGC) (White *et al.* 1990) were used to amplify the ITS region of the fungal rDNA, and primers bRPB2-6F (TGGGGYATGGTNTGYCCYGC) / bRPB2-7R (GAYTGRTTRTGRTCRGGGAAVGG) (Matheny 2005) were used to amplify a fragment of the *rpb2* gene (500-800 bp). As there are no commonly used universal primers for *rpb2*, we chose a pair that is assumed to be basidiomycete-specific to preferentially amplify basidiomycetes, which potentially have higher intragenome rDNA variation and dominate in the forest litter and soil of the study area (Baldrian *et al.* 2012). Each 50 µl reaction mixture contained 34.5 µl H_2O, 5 µl 10x buffer for DyNAzyme II

DNA polymerase, 3 µl 10 mg/ml BSA, 2 µl forward primer (final concentration, 10 pmol/µl), 2 µl reverse primer (final concentration, 10 pmol/µl), 1 µl template DNA, 1.5 µl 4% Pfu polymerase / DyNAzyme DNA Polymerase (final concentration, 0.06 U/µl), and 1 µl PCR Nucleotide Mix (10 mM). The program for PCR amplification reaction of fungal ITS consisted of initial denaturation at 94 ºC, 5 min, 20 cycles (94 ºC 30 sec, 51 ºC 45 sec, 72 ºC 90 sec) and a 15 min final extension at 72 ºC. For *rpb2* amplification, the program consisted of initial denaturation at 95 ºC, 1 min, 34 cycles (95 ºC 1 min, 61 ºC 1 min, 72 ºC 1 min) and a 10–min final extension at 72 ºC. Three parallel PCR reactions were run per sample. PCR products from the same sample were pooled and cleaned by using the Wizard SV Gel and PCR Clean–Up System (Promega, Madison, WI, USA). The concentration of the purified PCR product was quantified using the Quant–iT™ PicoGreen ds DNA kit (Invitrogen, Grand Island, NY, USA). For each sample, 500 ng of PCR product was again cleaned using the MinElute PCR Purification kit (Qiagen, Hilden, Germany). The tagging of individual samples was performed using the Roche MID adaptors and GS Junior Rapid Library Preparation Kit (Roche, Basel, Switzerland), following the manufacturer's instructions. The DNA content in each sample was quantified using the Kapa Library Quantification kit (KapaBiosystems, Woburn, MA, USA) following the manufacturer's instructions, and a equimolar mixture of all ITS samples and all *rpb2* samples was prepared. Pyrosequencing was run on the GS Junior (Roche, Basel, Switzerland).

Bioinformatic analysis

Pyrosequencing of the environmental samples yielded a total of 25,037 raw ITS forward sequences starting with the ITS1 primer and 26,712 raw *rpb2* forward sequences starting with the bRPB2-6F primer. The data were filtered and trimmed using the pipeline SEED (Větrovský & Baldrian 2013). All of the sequences with mismatches in tags were removed from the dataset. Pyrosequencing noise reduction was performed using the PyroNoise algorithm translation within Mothur (Schloss *et al.* 2009), chimeric sequences were detected using Uchime (Edgar *et al.* 2011) and deleted. A total of 11,866 ITS and 11,295 *rpb2* sequences were retained after the removal low-quality sequences (mean < 25), sequences shorter than 380 bases, and potentially chimeric sequences. Proportion of chimeric sequences was 24.3% in the ITS dataset and 0.3% in the *rpb2* dataset (Table S2, Supporting Information). The non-fungal sequences, i.e., those with a best hit in the

NCBI database (Benson *et al.* 2000) to a non-fungal taxon, were removed (105 reads in the ITS dataset; 6838 reads in the *rpb2* dataset). *rpb2* sequences were truncated to 400 bp and shorter than 400bp were removed, and the ITS sequences were truncated to contain the entire ITS1 and 5.8S rDNA regions using a fungal ITS extractor (Nilsson *et al.* 2009). Details of pre-processing workflow are shown in Table S2, Supporting Information. The dataset was randomly resampled at the same sampling depth of 4,457 sequences for both molecular targets prior to the diversity analyses. Double clustering (with a 98 and 97% similarity threshold (Lundberg *et al.* 2012)) was performed separately for the *rpb2* and ITS sequences to yield operational taxonomic units (OTUs) using Usearch (Edgar 2010). The consensus from each OTU was constructed from a MAFFT alignment (Katoh *et al.* 2009) based on the most abundant nucleotide at each position, an approach that enabled us to obtain sequences translatable into proteins. As the intraspecies variability of the *rpb2* gene should be similar to ITS (Schoch *et al.* 2012), we used the same similarity threshold for both genes. Additionally, we also compared the diversity and richness estimates for the dataset clustered on the level of 96, 97, and 98% sequence nucleotide similarity in SEED (Větrovský & Baldrian 2013).

Mock community composition and analysis

One hundred and thirty fungal species (83 from Basidiomycota, 42 from Ascomycota and 5 from other fungal groups), representing common wood and litter degrading and mycorrhizal fungi, were selected to generate an *in vitro* assembled mock community. DNA was isolated from axenic cultures or from freshly collected fruiting bodies (Table S3, Supporting Information) using ArchivePure DNA Yeast & Gram-+ Kit (5 PRIME, Hamburg). DNA isolated from individual strains was mixed in equimolar proportion. The pyrosequencing was performed using the same methods as in the soil samples. Data processing yielded in total 10 802 raw ITS sequences and 21 831 *rpb2* raw sequences. Those sequences in the *rpb2* dataset that were not attributable to this gene (6%) were excluded from analysis. After denoising, quality check and chimera removal, 7668 sequences remained in the ITS dataset and the *rpb2* dataset was resampled to the same size. Proportion of chimeric sequences was 7.4% in the ITS dataset and 1.3% in the *rpb2* dataset. Double clustering and further processing was the same as in environmental sequences (Table S2, Supporting Information).

OTU identification, phylogenetic analyses, and statistics

In soil datasets, NCBI tBlastX was used to generate the best species hits for the OTU consensus sequences of *rpb2*, and BlastN and the PlutoF pipeline (Tedersoo *et al.* 2010) were used in the case of the ITS data. Each OTU was assigned to the taxonomic level of class (or nearest lower or higher level when the class was not specified) using the taxonomic information from the NCBI taxonomy server. A tBlastX search using the *rpb2* gene often showed little similarity to identified sequences, preventing reliable identification (mostly in OTUs belonging to the basal fungal lineages). Thus, phylogeny-based taxonomic assignments of the *rpb2* OTUs were also performed. Lastly, the taxonomic spectra of the *rpb2* and ITS datasets were compared to identify differences in primer selectivity.

A data matrix for the phylogenetic analyses was constructed using a pooled dataset consisting of the consensus sequences of 340 *rpb2* OTUs, the alignment published by Lutzoni *et al.* (2004), and the best BLAST matches from the NCBI GenBank. Introns were detected based on the alignment obtained using the multiple alignment tool MAFFT server (http://mafft.cbrc.jp/alignment/server/) and the nucleotide sequences were translated into amino acid sequences. The sequences that showed errors (single-base indels), preventing their translation, were excluded from all of the analyses. The relatedness of the protein sequences was inferred using the Maximum likelihood phylogenetic analysis computed with the GTR substitutions model and the Fast likelihood-based method (aLTR SH-like) of branch support estimation in PhyMLOnline (http://atgc.lirmm.fr.phyml/; Guindon 2010). The tree was edited in FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree). The amino acid alignment consisted of 587 sequences in 146 positions of which 139 were variable and 117 parsimony informative. OTUs with an uncertain position within the basal fungal lineages were suspected to represent potential paralogues and pseudogenes. We confirmed their orthologous nature using a similarity search against the annotated Swiss-Prot database using a BlastX search in NCBI GenBank and on the basis of indices measuring codon usage bias, which were calculated using the methods described by Hubka & Kolařík (2012).

In the mock datasets, OTUs were identified using BlastN similarity search against the Genbank NCBI database and, where necessary, against local databases of sequences obtained by Sanger sequencing. Only species that were represented at least by 30 sequences in both datasets were used in comparison of intragenomic variability of the ITS

and *rpb2* genes and were both resampled at 30 sequences per species. OTU construction was performed as described above.

The software package STAMP (Parks & Beiko 2010) was used for the analysis and comparison of taxonomic unit profiles (Parks & Beiko 2010) using Fisher's exact test with a Bonferroni correction. Differences at $P < 0.01$ were regarded as statistically significant. A sign test was used for the comparison of diversity estimates using the ITS and *rpb2* markers in individual samples. Differences at $P < 0.05$ were regarded as statistically significant.

Results

Clustering of environmental sequences at a 97% similarity threshold yielded 497 OTUs in the ITS dataset (273 singletons, 55% of all OTUs, 6.1% of reads) and 340 OTUs in the *rpb2* dataset (149 singletons, 44% of all OTUs, 3.3% of all reads). A nucleotide sequence similarity ≥ 97% with the GenBank hits was found for 37% of ITS OTUs and for 2% of *rpb2* OTUs (Figure S1, Supporting Information).

The taxonomic assignments of the ITS and the *rpb2* sequences were compared at the level of class or neighbouring taxonomic units (Figure 1, Table 1). The relatively long and conserved partial *rpb2* protein sequence enabled the construction of a phylogenetic tree covering all fungi (Figure 2), which proved to be useful for the taxonomic placement of fungi within the underexplored basal lineages for which tBlastX did not provide reliable identification. The distribution of OTUs among 35 recognised fungal lineages (ITS, 21 lineages; *rpb2*, 28 lineages) (Figure 1, Table 1) suggests that both primer sets have broad taxonomic coverage. The non-fungal *rpb2* sequences (1799 OTUs) belonged to all major groups of typical soil organisms, including archaea, bacteria, various protists, soil fauna, algae, and plants (Table S4, Supporting Information). Although the ITS dataset contained more OTUs classified into Ascomycota (ITS: 59%, *rpb2*: 27% of all fungal OTUs) and a similar number assigned to Basidiomycota (ITS: 31%, *rpb2*: 32% of all fugal OTUs), the number of sequences assigned to basal fungal lineages was significantly higher in the *rpb2* dataset (ITS: 9%, *rpb2*: 41% of all OTUs). In both datasets, the phylum Basidiomycota was the most dominant, followed by Ascomycota and other fungi, when the number of sequence reads was taken into account.

The dominant part of the fungal communities, as indicated by the ITS marker, was represented by the groups of Agaricomycetes, Leotiomycetes, Eurotiomycetes, Dothideomycetes, Tremellomycetes, Pucciniomycotina, and Mortierellales (ordered based on sequence abundance). This result is in general agreement with the spectrum identified using tBlastX within the *rpb2* dataset (Table 2), with most sequences belonging to Agaricomycetes, Eurotiomycetes, a group of undefined basal lineages, Leotiomycetes, Dothideomycetes, Pucciniomycotina, and Mortierellales. From the more abundant major groups, only Glomeromycetes and Saccharomycetes were found in ITS but were absent or very rare in the *rpb2* dataset.

Although the spectrum of the most abundant OTUs identified by both primer sets was remarkably overlapping, their abundance differed largely. An example is the genus

Lactarius, which was represented by 40% of all sequences in the *rpb2* dataset, but only 1.3% sequences in the ITS dataset (Table 2). Conversely, *Piloderma*, which dominated among the ITS sequences (10.6%), was rare in the *rpb2* dataset (2.4%). The identified spectrum included symbionts of forest trees and ericoid, saprotrophic ascomycetes, basidiomycetes and zygomycetes constituting the core part of the fungal community.

In our study, there was high diversity and abundance of sequences classified into various groups of the basal fungal lineages in the *rpb2* dataset (41% of OTUs, 23% of reads). From these, the most abundant OTUs were clustered with *Rozella* (*Cryptomycota*), *Chytridiomycetes*, or belonged to several clades with uncertain position between the known lineages of basal fungi (Figure 2).

The two alternative markers were also used for the exploration of fungal richness. The OTU richness in individual samples was quantified at 209 sequences per sample and ranged from 34 to 80, with an average of 57 in the *ITS* dataset, which was significantly higher than in the *rpb2* dataset (26-66 OTUs, average of 46). The Chao1 estimates were also significantly higher for *ITS*, though differences in the Shannon-Wiener index and community evenness were not significantly different. Interestingly, the *rpb2* but not *ITS* analysis showed a high level of significance for differences in the species richness and diversity estimates between the litter and soil (Figure S2, Supporting Information). The OTU richness and diversity of the entire dataset was higher for the *ITS* dataset (497 OTUs, Shannon-Wiener Index of 4.45, Simpson diversity index of 0.031, and Chao1 of 940) than the *rpb2* dataset (340, 3.63, 0.085, and 527, respectively). The difference in Chao1 estimates was primarily due to the higher number of singletons in the *ITS* dataset (270 vs. 149) and the datasets with excluded singletons showed similar diversity estimates (ITS dataset: 227 OTUs, Shannon-Wiener Index of 4.13, Simpson diversity index of 0.035, and Chao1 of 227; *rpb2* dataset: 191 OTUs, Shannon-Wiener Index of 3.43, Simpson diversity index of 0.091, and Chao1 of 191). We observed a lower effect of nucleotide similarity on the resulting OTU numbers for *rpb2* than the ITS dataset: the *rpb2* dataset showed only a slight increase in OTU number among 90-97.5% similarity (Figure S3, Supporting Information). The similarity used in our study in both genes (98% pre-clustering, 97% clustering) roughly corresponds to the similarity of 96.5% applied to the original sequence set (without preclustering), showing that the effect of different interspecies variability in both genes does not significantly influence the estimated diversity.

The identification of *rpb2* OTUs using tBlastX was sufficiently straightforward in most cases when the sequence similarity was sufficiently high (99-85%). For those OTUs that did not have close matches in GenBank, their position within the fungal tree of life had to be inferred using a phylogenetic analysis of their protein sequences. The *rpb2* sequences were found to be conserved enough to allow the construction of a robust phylogenetic tree covering all of Eukaryota. The chosen *rpb2* marker region only rarely contains introns (detected only in three OTUs in this study), which is another advantage for phylogenetic comparisons. Amplified rbp2 region can contain up to ten introns as showed by our in silico analysis, but these are located out of the sequenced part, and only few fungi from basal lineages contained one or two introns (Table S5, Supporting Information).

Our tree was in general agreement with the fungal tree of life (Hibbett et al. 2007; Jones et al. 2011, James et al. 2006). Several OTUs were found to form a new lineage sister to the *Ustilaginomycotina*. Members of the *Chaetothyriales* were not sister with other members of *Eurotiomycetes*, as it is known from multigene phylogenies and were placed close to *Dothideomycetes* in agreement with morphology and *rpb2* based phylogenetic comparisons (Liu and Hall 2004). Three of the most dominant OTUs (Table 2) and several other OTUs belonged to lineages identical or close to the *Cryptomycota*. Other lineages of unknown identity are marked in Figure 1 and Table 1.

Interpretation of the previous results is limited by the fact that the real species diversity, taxon spectrum and abundance in the analysed soil community is not known. To overcome this limitation, mock community consisting of equimolar DNA concentrations of 130 fungal taxa was studied. Clustering resulted in 275 (120 singletons, 43 % of all OTUs, 1.5% of all reads) and 177 OTUs (84 singletons, 47% of all OTUs, 1 % of all reads) in the ITS and *rpb2* datasets, respectively. Sequences of all species were recovered using at least one marker: 68 were present in both datasets, whereas 29 and 33 species were only present in the *rpb2* and ITS datasets, respectively (Table S3, Supporting Information). All major fungal groups (*Ascomycota*, *Basidiomycota*, *Mortierellomycotina*, *Mucoromycotina*) were present in similar proportions in both datasets, when the number of species was considered. However, the observed diversity (OTU counts) and abundance (read counts) of these higher taxa in the ITS dataset differed significantly from expected values (X^2 test, p=0.036 for OTU number). The same comparison in *rpb2* dataset showed that observed OTU number differed from expected value (X^2 test, p=0.027) but number or reads did not (p=0.26). Most importantly, ITS highly overestimated the OTU counts and

relative abundance of the *Basidiomycota* (Table 3). Furthermore, considerable differences in taxon abundance between the two markers was observed which was also the case when comparing the soil community datasets. For example, only seven species ranked among the 20 most abundant species in both datasets.

OTU numbers estimated in the mock community exceeded the real number of species included with both ITS and *rpb2*. When the 15 species most abundant in both datasets were analysed at the same sequencing depth, ITS gave 2.53 ± 1.13 OTU per species and *rpb2* gave 1.87 ± 0.92 OTU per species, which was marginally less ($P = 0.083$).

Discussion

Universality of degenerated rpb2 primers

The *rpb2* primers used in this study were designated by their authors as basidiomycete-specific (Matheny 2005) and also our *in silico* analysis indicated their preference for Agaricomycetes and a lower specificity for other groups (Table S6, Supporting Information). Unexpectedly, in our study the *rpb2* gene produced a more taxonomically diverse set of fungal sequences than the universal *ITS* primers. The *rpb2* dataset contained much more non-fungal sequences than the ITS dataset (61% vs. 1% of reads). The non-fungal organisms amplified using the *ITS* primers primarily belonged to *Ciliata* and *Viridiplantae*, which is in line with the fact, that members of fungi, ciliates and plants were used by White et al. (1990) for ITS1 and ITS4 primer design and our data confirmed that these primers are biased towards these eukaryotic groups (Table S4, Supporting Information). The provenience of nonfungal sequences in the *rpb2* dataset was much complex, containing all major groups of soil biota. In *ITS* and *rpb2* datasets from environmental samples a proportionally similar number of basidiomycete reads (ITS: 51%, *rpb2*: 58% of all reads) was found in both datasets, showing that both primer sets had similar taxonomic coverage in *Dikarya* (*Ascomycota* and *Basidiomycota*). The broad taxonomic coverage of *rpb2* primers was fully supported by the results of the mock community analysis. It is widely known that use of degenerate primers can greatly reduce the specificity of PCR amplifications, though the mechanisms responsible for such universality of degenerate *rpb2* primers in the amplification of a highly complex DNA template mixture is unknown. The existence of universal primers with sufficient taxonomic resolution, is one of the major criteria required for a universal barcode in fungi (Schoch et al. 2012). Single-copy protein-coding genes provide good taxonomic resolution but are considered inaccessible using universal primers (Schoch et al. 2012). Here we show that lower primer universality is not necessarily a limitation for the metabarcoding.

Comparison of the ITS and rpb2 datasets at various taxonomic levels

The fungal taxon spectra obtained using the *rbp2* and *ITS* markers corresponded both in the environmental and mock communities, but taxon abundances differed widely. The soil and litter of the coniferous forests of the boreal and temperate zones are considered to be

dominated by saprotrophic and mycorrhizal *Dikarya*, which represented typically about 90% of OTUs in the ITS datasets (Buéé et al. 2009, O'Brien et al. 2005, Baldrian et al. 2012) and also in our study. The most notable difference between the *ITS* and *rpb2* datasets from our study was the high diversity and abundance of sequences classified into various groups of the basal fungal lineages in the *rpb2* dataset (41% of OTUs, 23% of reads). Basal fungal lineages are often zoosporic, difficult to cultivate, widespread in water ecosystems or soil; their members live as saprobes, symbionts, or parasites (Jones et al. 2011; Marano et al. 2012). Freeman (2009) found that Chytridiomycota constituted over 70% of rDNA sequences in high-elevation soils without vegetation cover. High abundance of chytrids (10% of sequences) was also reported from periodically flooded alpine tundra soil under the snow cover (Freeman 2009). Similar conditions, where zoosporic fungi can thrive, occur in the forest from the present study and thus they may represent an abundant part of the mycobiota. It is possible that zoosporic fungi were abundant in the studied ecosystem and yet were underestimated by the *ITS* marker due to the lack of complementarity of *ITS* primers or possibly due to the proportionally lower rDNA copy numbers per genome in these fungi. Future research should address whether these lesser known fungi constitute a significant portion of the mycobiome of coniferous forests.

Diversity estimation using ITS and rpb2 gene

The published fungal diversity estimates from environmental samples based on ITS analyses are notoriously high (Blackwell 2011) and should be critically re-evaluated. The presence of chimeras, deeper paralogues in multicopy markers and pseudogenes is the main source of error. In our environmental samples and mock community, the number of singletons was much higher in the ITS dataset. The same was observed for the OTU diversity in environmental samples. This is evidently attributable to the higher intragenic variability of the ITS marker as shown by our comparative analysis of the mock community. The ITS dataset may theoretically contain deeper intragenomic paralogues, which increased number of singletons and OTUs a in the *ITS* dataset. The difference in the spectrum of fungal groups recovered by the two primer sets may be an alternative explanation because the degree of intragenomic and intraspecies variability is not equal across the fungal kingdom (Nilsson et al. 2008).

Our study revealed that multiple sequence clusters phylogenetically distant from known taxa were present in the environmental sample. Such lineages were also reported in several other studies. The one of Jumpponen (2007) indicated that such clusters located at more basal positions in the rDNA phylograms could be chimeric and that the proportion of such chimeric sequences could be very high. The fact that only DNA sequences that are translatable into protein were used in our phylogenetic analysis partly eliminates the risk of chimeras or pseudogenes with the protein-encoding markers. The *ITS* marker seemed to be much more sensitive to chimera formation than *rpb2* (proportion of chimeric sequences detected by UChime was 25.3% in ITS and 0.30% in *rpb2* sequences, Table S2, Supporting Information). Although there is the risk that unrecognised paralogues are present in the *rpb2* dataset, such as, for example, in another protein-coding gene, β-tubulin (Hubka *&* Kolarik 2012), genes paralogous to *rpb2* have not been reported thus far. The probability that the phylogenetically distant sequence clusters recovered using *rpb2* represent such paralogues is thus low and it is possible that these sequences correspond to already known fungal groups, whose *rpb2* sequences are missing in public databases. The unassigned lineages may also represent novel ones that were not previously amplified using *ITS* primers. The advantages of degenerate primers in the discovery of new uncultured fungi should thus be further evaluated.

Since similar levels of OTU inflation was observed for both markers in the soil community as in the defined mock community, this may suggest that the real diversity in the soil is closer to the estimate obtained using the *rpb2*. This inflation is partially attributable to higher intragenomic variability of the ITS region. The polymorphism was found mostly in the *Polyporales* and some members of the *Agaricales*. DNA in our study originated from monokaryotic haploid *Ascomycota* cultures as well as dikaryotic cultures or fruiting bodies of the *Basidiomycota*. The heterokaryotic members of the latter group can theoretically possess divergent alleles that may cause the presence of two separate OTUs after clustering and it seems probable that intraspecies variability in both markers may affect diversity estimates in real samples. It should be also noted that the distribution of sequences among taxa of the mock community was slightly more even in the *rpb2* dataset where 80% sequences were represented by 12 species compared to 7 species in the ITS. Still, the reason for highly uneven distribution of read counts among species that was also observed in previous studies (Ihrmark *et al.* 2012), remains unclear: theoretically, it can be due to the combination of primer specificity, PCR preference, DNA quality, or other factors.

rpb2 as an alternative metabarcoding marker

The ITS gene currently offers an unmatched opportunity for the close identification of a particular OTU. However, the present study shows that the single-copy, protein-encoding gene *rpb2* may be a viable option for fungal metabarcoding. Our results show general agreement in the identity of the fungal genera and fungal classes recovered using this marker and *ITS*. No such agreement was found in the sequence abundance of the main genera or major fungal groups. Indeed, the single-copy nature of *rpb2* represents an important advantage for proper estimations of diversity and relative abundance; furthermore, the constant length of the *rpb2* amplicon avoids the PCR bias observed in the case of *ITS*, where length varies largely among taxa (Figure S4, Supporting Information). The *rpb2* gene possesses a barcode gap between the inter- and intraspecific variation that is much more clearly defined than in the ITS1 sequence, which corresponds to the results of Schoch *et al.* (2012). The *rpb2* gene have taxonomic sensitivity superior to the *ITS* (Schoch *et al.* 2012) and our results reveal that this sequence is well suited for the study of basal fungal lineages. The use of a translatable protein-coding gene also enables the identification of potential pseudogenes and the construction of robust phylogenetical trees. Although the precise taxonomic placement of *rbp2* sequences is currently more difficult than it is for *ITS* due to the lower representation in GenBank, the phylogenetic discriminative power, better quantitative representation of the community composition, and suitability for phylogenetic analyses may represent comparative advantages for *rbp2* over the use of *ITS* and make this molecular marker useful for studies in fungal ecology and diversity.

Acknowledgements

This work was supported by the Czech Science Foundation (13-06763S, P504/12/0709), the Ministry of Education, Youth and Sports of the Czech Republic (LD12050), and by the Research concept of the Institute of Microbiology ASCR (RVO61388971).

References

Baldrian P, Kolařík M, Štursová M *et al.* (2012) Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME Journal*, **6**, 248-258.

Baldrian P, Větrovský T, Cajthaml T *et al.* (2013) Estimation of fungal biomass in forest litter and soil. *Fungal Ecology*, **6**, 1-11.

Bellemain E, Carlsen T, Brochmann C *et al.* (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiology*, **10**, 189.

Benson DA, Karsch-Mizrachi I, Lipman DJ *et al.* (2000) GenBank. *Nucleic Acids Research*, **28**, 15-18.

Blackwell M (2011) The fungi: 1, 2, 3 ... 5.1 million species? *American Journal of Botany*, **98**, 426-438.

Buéé M, Reich M, Murat C *et al.* (2009) 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist*, **184**, 449-456.

Connell LB, Redman R, Rodriguez R *et al.* (2010) *Dioszegia antarctica* sp. nov. and *Dioszegia cryoxerica* sp. nov., psychrophilic basidiomycetous yeasts from polar desert soils in Antarctica. *International Journal for Systematic and Evolutionary Microbiology*, **60**, 1466-1472.

Debode J, Van Hemelrijck W, Baeyen S *et al.* (2009) Quantitative detection and monitoring of *Colletotrichum acutatum* in strawberry leaves using real-time PCR. *Plant Pathology*, **58**, 504-514.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194-2200.

Engelbrektson A, Kunin V, Wrighton KC *et al.* (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME Journal*, **4**, 642-647

Feau N, Decourcelle T, Husson C, Desprez-Loustau ML, Dutech C (2011) Finding single copy genes out of sequenced genomes for multilocus phylogenetics in non-model fungi. *PLoS ONE*, **6**, e18803.

Freeman KR, Martin AP, Karki D *et al.* (2009) Evidence that chytrids dominate fungal communities in high-elevation soils. *Proceedings of the National Academy of Sciences of the USA*, **106**, 18315-18320

Gihring TM, Green SJ, Schadt CW (2012) Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environmental Microbiology*, **14**, 285-290.

Glass DJ, Takebayashi N, Olson LE, Taylor DL (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**, 234-245.

Guindon S, Dufayard JF, Lefort V *et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307-321.

Herrera ML, Vallor AC, Gelfond JA, Patterson TF, Wickes BL (2009) Strain-dependent variation in 18S ribosomal DNA copy numbers in *Aspergillus fumigatus*. *Journal of Clinical Microbiology*, **47**, 1325-1332.

Hibbett DS, Binder M, Bischoff JF *et al.* (2007) A higher-level phylogenetic classification of the Fungi. *Mycological Research*, **111**, 509-547.

Hubka V, Kolařík M (2012) β-tubulin paralogue tubC is frequently misidentified as the benA gene in *Aspergillus* section *Nigri* taxonomy: primer specificity testing and taxonomic consequences. *Persoonia*, **29**, 1-10.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*, **12**, 1889-1898.

Chao A (1984) Non-parametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, **11**, 265-270.

Ihrmark K, Bödeker IT, Cruz-Martinez K *et al.* (2012) New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology*, **82**, 666-677.

James TY, Kauff F, Schoch CL *et al.* (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, **443**, 818-822.

Jones MDM, Forn I, Gadelha C *et al.* (2011) Discovery of novel intermediate forms redefines the fungal tree of life. *Nature*, **474**, 200-203.

Jumpponen A (2007) Soil fungal communities underneath willow canopies on a primary successional glacier forefront: rDNA sequence results can be affected by primer selection and chimeric data. *Microbial Ecology*, **53**, 233-246.

Katoh K, Asimenos G, Toh H (2009) Multiple Alignment of DNA Sequences with MAFFT. In: Posada D (ed) *Bioinformatics for DNA Sequence Analysis*. Humana Press Inc., Totowa, pp 39-46.

Kåuserud H, Schumacher T (2003) Ribosomal DNA variation, recombination and inheritance in the basidiomycete *Trichaptum abietinum*: Implications for reticulate evolution. *Heredity*, **91**, 163-172.

Kiss L (2012) Limits of nuclear ribosomal DNA internal transcribed spacer (ITS) sequences as species barcodes for Fungi. *Proceedings of the National Academy of Sciences of the USA*, **109**, E1811.

Keeling PJ, Inagaki Y (2004) A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1 alpha. *Proceedings of the National Academy of Sciences of the USA*, **101**, 15380-15385.

Kruger M, Kruger C, Walker C, Stockinger H, Schussler A (2012) Phylogenetic reference data for systematics and phylotaxonomy of arbuscular mycorrhizal fungi from phylum to species level. *New Phytologist*, **193**, 970-984.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology*, **12**, 118-123.

Lindahl BD, Nilsson RH, Tedersoo L *et al.* (2013) Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New Phytologist*, **199**, 288-299.

Lindner DL, Banik MT (2011) Intragenomic variation in the ITS rDNA region obscures phylogenetic relationships and inflates estimates of operational taxonomic units in genus *Laetiporus*. *Mycologia*, **103**, 731-740.

Lindner DL, Carlsen T, Nilsson RH, *et al.* (2013) Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution*, **3**, 1751-1764.

Liu YJ, Hall BD (2004) Body plan evolution of ascomycetes, as inferred from an RNA polymerase II, phylogeny. *Proceedings of the National Academy of Sciences of the USA*, **101**, 4507-4512.

Lucking R, Lawrey JD, Gillevet PM *et al.* (2014) Multiple ITS haplotypes in the genome of the lichenized basidiomycete *Cora inversa* (Hygrophoraceae): Fact or artifact? *Journal of Molecular Evolution*, **78**, 148-162.

Lundberg DS, Lebeis SL, Paredes SH *et al.* (2012) Defining the core *Arabidopsis thaliana* root microbiome. *Nature,* **488**, 86-90

Lutzoni F, Kauff F, Cox CJ *et al.* (2004) Assembling the fungal tree of life: Progress, classification and evolution of subcellular traits. *American Journal of Botany*, **91**, 1446-1480.

Marano AV, Gleason FH, Bärlocher F *et al.* (2012) Quantitative methods for the analysis of zoosporic fungi. *Journal of Microbiological Methods*, **89**, 22-32.

Matheny PB (2005) Improving phylogenetic inference of mushrooms with RPB1 and RPB2 nucleotide sequences (*Inocybe*; *Agaricales*). *Molecular Phylogenetics and Evolution*, **35**, 1-20.

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson KH (2008) Intraspecific ITS variability in the Kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary Bioinformatics*, **4**, 193-201.

Nilsson RH, Ryberg M, Abarenkov K, Sjokvist E, Kristiansson E (2009) The ITS region as a target for characterization of fungal communities using emerging sequencing technologies. *FEMS Microbiology Letters*, **296**, 97-101.

O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M, Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology*, **71**, 5544-5550.

Öpik M, Metsis M, Daniell T, Zobel M, Moora M (2009) Large-scale parallel 454 sequencing reveals host ecological group specificity of arbuscular mycorrhizal fungi in a boreonemoral forest. *New Phytologist*, **184**, 424-437.

Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. *Bioinformatics*, **26**, 715-721.

Quince C, Curtis TP, Sloan WT (2008) The rational exploration of microbial diversity. *ISME Journal*, **2**, 997-1006.

Raidl S, Bonfigli R, Agerer R (2005) Calibration of quantitative real-time Taqman PCR by correlation with hyphal biomass and its copies in mycelia of *Piloderma croceum*. *Plant Biology*, **7**, 713-717.

Sagova-Mareckova M, Cermak L, Novotna J *et al.* (2008) Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Applied and Environmental Microbiology*, **74**, 2902-2907.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537-7541.

Schloss PD (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology*, **6**, e1000844.

Schoch CL, Seifert KA, Huhndorf S *et al.* (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 USA*, **109**, 6241-6246.

Simon UK, Weiss M (2008) Intragenomic variation of fungal ribosomal genes is higher than previously thought. *Molecular Phylogenetics and Evolution*, **25**, 2251-2254.

Smith ME, Douhan GW, Rizzo DM (2007) Intra-specific and intra-sporocarp ITS variation of ectomycorrhizal fungi as assessed by rDNA sequencing of sporocarps and pooled ectomycorrhizal roots from a *Quercus* woodland. *Mycorrhiza*, **18**, 15-22.

Stockinger H, Krueger M, Schuessler A (2010) DNA barcoding of arbuscular mycorrhizal fungi. *New Phytologist*, **187**, 461-474.

Tedersoo L, Nilsson RH, Abarenkov K *et al.* 2010. 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist* **188**, 291-301.

Toju H, Tanabe AS, Yamamoto S, Sato H (2012) High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples. *PLoS ONE*, **7**, e40863.

Větrovský T, Baldrian P (2013) Analysis of soil fungal communities by amplicon pyrosequencing: current approaches to data analysis and the introduction of the pipeline SEED. *Biology and Fertility of Soils*, **49**, 1027-1037.

Voříšková J, Baldrian P (2013) Fungal community on decomposing leaf litter undergoes rapid successional changes. *ISME Journal*, **7**, 477-486.

Wang D-M, Yao Y-J (2005) Intrastrain internal transcribed spacer heterogeneity in *Ganoderma* species. *Canadian Journal of Microbiology*, **51**, 113-121.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols: A guide to methods and applications*, Academic Press, pp. 315-322.

Youssef N, Sheik CS, Krumholz LR *et al.* (2009) Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Applied and Environmental Microbiology*, **75**, 5227-5236.

Data Accessiblity

Sequence data have been deposited in the MG-RAST public database (http://metagenomics.anl.gov/, dataset numbers 4538645.3 for the fungal ITS region and 4538646.3 for the *rpb2* gene).

Author Contributions

T.V., M.K. and P.B. designed and conducted the research. L.Z., P.B. and T.V. collected soils, extracted DNA and analysed the data, T.V., M.K. and **T.Z. constructed and analysed the mock community**. M.K., T.V. and P.B. wrote the first draft of the manuscript. All authors edited the manuscript.

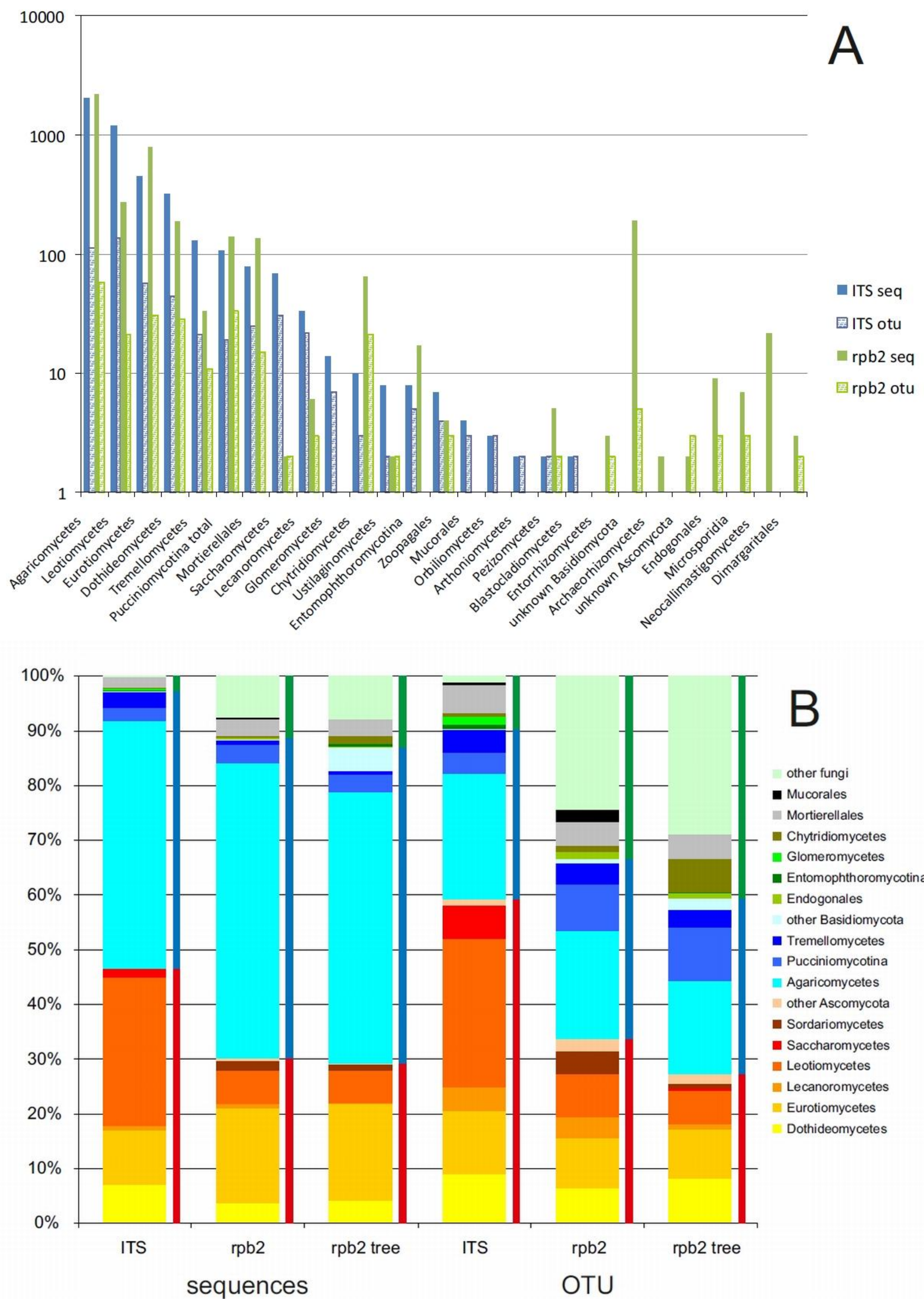

Figure 1. Abundance of sequences and OTUs at the level of fungal class identified from the ITS and *rpb2* datasets. A – Abundances of fungal classes, B – Relative composition of fungal communities based on sequence and OUT counts. The thin bars in the panel B use red color to indicate the Ascomycota, blue for the Basidiomycota and green for other fungal lineages.

Figure 2. Maximum likelihood phylogenetic tree showing the relationships of *rpb2* OTUs, the most related GenBank entries retrieved using tBlastX, and other representatives of major fungal groups. The clades containing OTUs only were collapsed, and the number of OTUs (O) and sequences (S) is marked.

Table 1. Relative abundance of sequences and OTUs at the level of fungal class identified from the ITS and rpb2 in spruce forest topsoil. The asterisks indicate significant differences in the relative abundance of sequences and OTUs based on ITS BlastN and rpb2 tBlastX, $^{*}P < 0.01$, $^{**}P < 0.001$.

	Fungal class	ITS BLASTn		RPB2 tBLASTx		RPB2 tree	
		reads	OTUs	reads	OTUs	reads	OTUs
Basidiomycota	Agaricomycetes	2015**	113	2409	67	2207	58
	Tremellomycetes	127**	21	43	13	33	11
	Agaricostilbomycetes	57**	10	6	1		
	Cystobasidiomycetes	0	0	10	2	138	33
	Microbotryomycetes	24**	6**	127	26		
	Pucciniomycetes	27**	3	0	0		
	Exobasidiomycetes	0	0	3	2	0	0
	Ustilaginomycetes	8	1	0	0	1	1
	Entorrhizomycetes	0	0	3	1	3	1
	unknown Basidiomycota	-	-	-	-	194	5
	Basidiomycota total	**2258****	**154**	**2601**	**112**	**2576**	**109**
Ascomycota	Arthoniomycetes	1	1	0	0	0	0
	Dothideomycetes	316**	45	167	22	186	28
	Eurotiomycetes	446**	57	769	31	785	31
	Lecanoromycetes	33	22	33	13	6	3
	Leotiomycetes	1209**	134**	282	27	269	21
	Lichinomycetes	0	0	5	2	0	0
	Orbiliomycetes	3	3	2	1	0	0
	Pezizomycetes	2	2	0	0	5	2
	Archaeorhizomycetes	0	0	2	1	2	1
	Pneumocystidomycetes	0	0	3	1	0	0
	Schizosaccharomycetes	0	0	3	3	0	0
	Sordariomycetes	0**	0**	76	14	52	3
	Saccharomycetes	68**	31**	0	0	1	1
	unknown Ascomycota	-	-	-	-	2	3
	Ascomycota total	**2078****	**295****	**1342**	**115**	**1308**	**93**
Basal lineages	Blastocladiomycetes	1	1	1	1	0	0
	Entomophthoromycotina	8	5	0	0	17	1
	Endogonales	0	0	10	4	9	3
	Mucorales	4	3	12	7	0	0
	Chytridiomycetes	10	3	16	4	64	21
	Microsporidia	0	0	7	3	7	3
	Monoblepharidomycetes	0	0	3	1	0	0
	Mortierellales	77**	25	140	15	137	15
	Glomeromycetes	14**	7	0	0	0	0
	Neocallimastigomycetes	0**	0	22	1	22	1
	Dimargaritales	0	0	0	0	3	2
	Zoopagales	7	4	6	3	4	3
	basal lineages of uncertain position	0**	0**	297	74	310	89
	Basal lineages total	**121****	**48****	**514**	**113**	**573**	**138**
	Total	4457	497	4457	340	4457	340

Table 2. Identification of the thirty most abundant OTUs in spruce forest topsoil using ITS and *rpb2* as alternative molecular markers. The identity was assigned using BlastN (for ITS) or tBlastX (for *rpb2*). The percent similarity of protein sequence is given in the case of *rpb2*. The taxonomic classification to *Ascomycota* (A) and *Basidiomycota* (B) is marked.

ITS					***rpb2***				
Identification	Accession number	Similarity [%]	Coverage [%]	Abundance [%]	Identification	Accession number	Similarity [%]	Coverage [%]	Abundance [%]
Piloderma (B)	JQ711958	97	100	10.6	*Lactarius* (B)	DQ408128	98	99	22.9
Lachnellula (A)	KC464638	98	100	7.0	*Lactarius* (B)	DQ408128	98	99	13.3
Mycena (B)	EF093152	100	100	6.2	*Ceramothyrium* (A)	AY485617	79	99	8.2
Cenococcum (A)	AM087244	99	100	4.4	*Rasamsonia* (A)	JQ729684	97	99	4.3
Marasmius (B)	FR717227	99	97	4.1	*Lactarius* (B)	DQ408128	99	99	3.6
Cladophialophora (A)	EF016381	97	100	3.8	*Hydropus* (B)	DQ472722	71	62	3.5
Russula (B)	HM189931	99	100	2.9	*Albotricha* sp. (A)	AB481347	98	99	3.2
Hygrophorus (B)	JF908073	100	100	2.5	*Piloderma* (B)	GU187797	99	99	2.4
Meliniomyces (A)	HQ157837	100	100	2.4	*Ceramothyrium* (A)	AY485617	92	90	2.3
Tylospora (B)	JN943896	100	100	2.4	*Gymnopus* (B)	DQ472716	99	97	1.7
Tricholoma (B)	AB036899	99	100	1.8	*Cudonia* (A)	AY641033	92	99	1.4
Uncultured (*Trechispora*) (B)	FJ475683	99	100	1.8	*Tyrannosorus* (A)	DQ470928	74	52	1.4
Cladophialophora (A)	EF016381	96	100	1.5	*Veluticeps* (B)	HM536125	92	100	≤1
Ascomycete (*Scytalidium*) (A)	GU067746	99	100	1.5	*Umbelopsis**	DQ302787	64	96	≤1
Inocybe (B)	AJ889955	99	99	1.5	*Ellisembia* (A)	DQ435090	79	94	≤1
Lactarius (B)	JQ712010	99	100	1.3	*Microbotryum* (B)	DQ789985	90	99	≤1
Rhizoscyphus (A)	JQ711796	96	100	1.3	*Mortierella*	DQ302784	92	100	≤1
Lyophyllum (B)	HE819396	99	45	1.3	*Microthyrium* (A)	GU371734	77	99	≤1
Lachnellula (A)	AB481245	98	100	1.3	*Microbotryum* (B)	DQ789985	90	99	≤1
Xerocomus (B)	HQ207696	100	100	1.1	*Spiromyces**	DQ302790	54	99	≤1
Cadophora (A)	AB543058	97	87	≤1	*Microthyrium* (A)	GU371734	75	99	≤1
Rhizoscyphus (A)	JQ711893	95	100	≤1	*Hydropus* (B)	DQ472722	71	62	≤1
Mortierella	JQ272448	99	100	≤1	*Mortierella**	DQ302784	95	99	≤1
Amanita (B)	EF493271	100	100	≤1	*Mortierella**	DQ302784	78	96	≤1

Meliniomyces (A)	HQ157926	99	100	≤1	*Ceramothyrium* (A)	AY485617	89	99	≤1
Leptodontidium (A)	GU067735	96	100	≤1	*Mortierella**	DQ302784	92	86	≤1
Exophiala (A)	HE605215	87	100	≤1	*Trichopeziza* (A)	AB481360	98	98	≤1
Tylospora (B)	JN943896	99	100	≤1	*Coprinopsis* (B)	XM1829088	100	93	≤1
Uncultured (*Leptodontidium*)	HM488455	100	100	≤1	*Gymnopus* (B)	DQ472716	99	99	≤1
Uncultured (*Rhodosporidium*)	FJ475820	100	62	≤1	*Mortierella**	DQ302784	92	92	≤1

*OTUs with low similarity to known fungal sequences were placed into basal fungal lineages of uncertain position based on the phylogenetic tree.

Table 3. Comparison of the representation of major fungal groups in the mock community and in the sequence datasets obtained with ITS and *rpb2* as molecular markers.

	Mock community	ITS			*rpb2*		
		Species detected	OTUs	Reads (%)	Species detected	OTUs	Reads (%)
Basidiomycota	83 (63.8%)	64 (63.4%)	205 (74.5%)	90.0	68 (70.1%)	133 (75.1%)	62.9
Ascomycota	42 (32.3%)	32 (31.7%)	62 (22.5%)	8.6	25 (25.8%)	38 (21.5%)	35.3
Mucoromycotina and Mortierellomycotina	5 (3.9%)	5 (4.9%)	8 (3.0%)	1.4	4 (4.1%)	6 (3.4%)	1.8