Metabarcoding of insect-associated fungal communities: a comparison of internal transcribed spacer (ITS) and large-subunit (LSU) rRNA markers

Angelina Ceballos-Escalera[1,2], **John Richards**[1,2], **Maria Belen Arias**[1,2], **Daegan J. G. Inward**[3], **Alfried P. Vogler**[1,2]

[1]*Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK.*

[2]*Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK.*

[3]*Forest Research, Alice Holt Research Station, Farnham, Surrey, GU10 4LH, UK*

*Corresponding author: **Angelina Ceballos-Escalera**

Abstract

Complex communities of fungi are regularly characterised by deep sequencing with standard barcode markers, especially the Internal Transcribed Spacer (ITS) and the Large-Subunit (LSU) gene of the rRNA locus. Full taxonomic characterisation of fungal communities is necessary for establishing ecological associations or for early detection of pathogens and invasive species, but reliance on a single short sequence fragment may be problematic due to various technological and analytical obstacles. Here we conducted a side-by-side comparison of fungal communities associated with bark beetles (Scolytinae), the likely vectors of several tree pathogens, using ITS and LSU. Both markers revealed similar patterns of overall species richness and response to key variables (beetle species, forest type). Identification against the respective reference databases using various classifiers revealed similar higher-level taxonomic composition, but decreasing resolution towards lower levels, especially at the species level. Thus, Operational Taxonomic Units (OTUs) could not be linked via taxonomic classifiers across ITS and LSU fragments. However, a phylogenetic analysis (focused on the epidemiologically important Sordariomycetes) placed OTUs relative to reference sequences spanning both loci and demonstrated the largely similar phylogenetic distribution of ITS and LSU-derived OTUs. The analysis of congruence in both markers also suggested the biologically most defensible threshold values for OTU delimitation (98% for ITS, 99% for LSU). Studies of complex fungal communities using the canonical ITS metabarcode marker require corroboration across additional loci, and benefit from phylogenetic analyses and their greater precision compared to conventional taxonomic classifiers, even in the face of incomplete and partially identified reference databases.

Introduction

Fungal communities associated with insects have been widely studied to disentangle the nature of their interactions (Ganter 2006, Raman et al. 2012, Li et al. 2016, Malacrinò et al. 2017). For these studies to succeed, an accurate and reliable fungal identification is essential. However, these identifications are challenging due to the cryptic nature and incomplete taxonomy of the fungi, as only 3-8% of fungal species have been described so far (Hibbett et al. 2016, Kandawatte Wedaralalage et al. 2020). Conventional studies of fungal communities have been conducted by isolating and culturing each fungus recovered from the substrate (Batra 1963), which generated biases towards culturable species. High-throughput sequencing has provided an alternative methodology, especially through metabarcoding, i.e. the sequencing of short amplicons from mixed communities (Yu et al. 2012). Metabarcoding is now widely applied in characterising species composition and biodiversity of fungal

communities associated with insects. In the specific case of fungal communities associated with bark beetles, metabarcoding usually detects dozens of species of fungi on a single insect specimen (Bálint et al. 2014, Miller et al. 2016, 2019, Malacrinò et al. 2017, Johnson et al. 2018, Hulcr et al. 2020).

There is broad agreement that the internal transcribed spacer (ITS) of the nuclear rRNA gene cluster should be the standard DNA barcode in fungi (Schoch et al. 2012). Its use in metabarcoding is now equally well established, and extensive reference databases and universal primer combinations are in wide use (Porras-Alfaro et al. 2014, Tedersoo et al. 2015b). However, various challenges remain for accurate characterisation of communities. PCR amplification biases may skew species recovery (Bellemain et al. 2010, Harrington et al. 2011, Dreaden et al. 2014, Li et al. 2020), in particular given that primer design is constrained in studies of associations with insects, which requires to avoid co-amplification of the corresponding genomic regions of the hosts. In addition, the recovered short sequence fragments have limited power for placement in the phylogenetic system (Vrålstad 2011, Porras-Alfaro et al. 2014), exacerbated by the taxonomic insufficiencies of the reference databases (Porras-Alfaro et al. 2014, Tedersoo et al. 2015a, Miller et al. 2016, Agerbo Rasmussen et al. 2020). In response to these concerns many fungal phylogenetic and barcoding studies have used a combination of ITS and partial large and small subunit (LSU and SSU) rRNA genes, as well as other markers such as RPB2 and TEF1α (Lutzoni et al. 2004, Zhang et al. 2006, Stielow et al. 2015). The latter two markers lack representation in databases, while curated reference databases and analysis tools like SILVA and RDP (Ribosomal Database Project) are built specifically for SSU and LSU genes (Wang et al. 2007, Quast et al. 2012) .

In practise, both ITS and the LSU/SSU markers exhibit particularities whose benefits and drawbacks depend on the aim and scope of a study (Porras-Alfaro et al. 2014). While the LSU/SSU genes are less variable than the ITS intergenic region, which favours alignment and tree-based analyses, their low molecular rate reduces the taxonomic resolution at the species-level. In turn, ITS provides better species resolution, but the higher mutation rate leads to intragenomic variation of the tandem repeat units, and high variability often translates into poor alignments, making it difficult to apply phylogenetic methods (Vrålstad 2011, Porras-Alfaro et al. 2014). However, this intraspecific and intragenomic variation is still poorly documented, and may affect not only the ITS (Lücking et al. 2020). Furthermore, the suitability of the primers used to amplify fungal communities is often tested by using mock communities (Bakker 2018, Egan et al. 2018, Frau et al. 2019), but only a few studies have addressed the relative power of ITS and LSU/SSU markers in complex natural fungal communities (Parada et al. 2016, Jusino et al. 2019, Li et al. 2020).

A limited number of studies comparing different regions of the nuclear rRNA cluster have revealed similar ecological patterns and broad conclusions that are not profoundly affected by marker choice (Tedersoo et al. 2015b, George et al. 2019, Nilsson et al. 2019). However, these studies generally have applied a coarse-grain approach of higher-level taxonomic analysis, rather than the species level. As the databases become more complete and the quality of HTS reads is improving, increasingly precise matches of database and query allow a greater detail of community analysis, ultimately at the species (Operational Taxonomic Unit, OTU) or haplotype (Amplified Sequence Variant, ASV) level. In the few cases where this approach has been taken to date, the choice of markers seems to be critical. For example, the ITS marker may not detect key pathogen species that are recovered using LSU (Skelton et al. 2019, Hulcr et al. 2020). Furthermore, ideas of what constitutes a species and what are the most appropriate ways to circumscribe them based on barcode sequences are always shifting (Hyde et al. 2013, Kõljalg et al. 2013, Hibbett et al. 2016, Kandawatte Wedaralalage et al. 2020, Lücking et al. 2020).

This leaves the question about which marker is preferrable, and given the presumed advantages of each, to what degree does the marker choice matter for the conclusions from metabarcoding studies? Here we address these questions for the case of fungal communities associated with bark beetles (Coleoptera: Scolytinae). These insects breed in living or dead trees and form close associations with fungi important for access to nutrients from wood that cannot be utilised directly by the beetles themselves (Batra 1963). Fungal communities associated with these beetles are highly diverse and form symbioses of various strength and specificity, which in many cases involve the active transport of fungal associates in specifically adapted pockets of the beetles' exoskeleton, the mycangia (Six 2020). A beetle-fungus complex can cause enormous damage to forest ecosystems, e.g. resulting in the demise of chestnuts in North America, elms across the Northern Hemisphere, or the recent large-scale decline of conifer forests in Central Europe and North America, which usually involve fungi from the ascomycote orders Ophiostomatales, Microascales and Hypocreales (Class Sordariomycetes) (Ploetz et al. 2013). Metabarcoding provides a tool for detailed studies of these complex communities, but the results may be influenced by the choice of barcode markers and other problems of using short sequences from PCR mixtures to characterise complex species communities. Confidence in the metabarcoding approach can be improved by analysing ITS and LSU marker sets from a given sample in parallel, which should produce the same outcome. However, given different evolutionary rates, the limited power of phylogenetic placement, and different resolution of species differentiation may result in mismatched conclusions from either marker. We used individuals from four bark beetle species obtained from three forest types to characterise the associated fungal communities, conducting a

comparison of the two markers with regard to: (1) broad ecological trends of fungal associations taking a whole-community approach, and (2) species identifications against existing ITS and LSU fungal reference databases, using various taxonomic classifiers and explicit phylogenetic methods. The side-by-side comparison addresses the power of either markers to infer critical parameters of fungal community metabarcoding, such as the number and taxonomic identity of OTUs, their ecological associations, and the inferences on whole-community diversity and turnover.

Materials and Methods

Samples used and laboratory procedures

Sequence data were generated from 20 specimens each of four beetle species: *Xylosandrus germanus*, *Xyleborinus saxesenii*, *Gnathotrichus materiarius*, and *Tomicus piniperda* (**Table 1**). Only the latter is a xylophagous 'bark beetle' in the strict sense, while the three others are considered mycelia feeding (xylomycetophagous) 'ambrosia beetles' that apparently rely on active transport of fungi indicated by the presence mycangia (see Six, 2020). Specimens were collected by Forest Research during 2013-2015 in the New Forest National Park (50º50'52.08" N 1º35'33.51" W), Hampshire, UK, using Lindgren multiple-funnel traps (Lindgren 1983) (Phero Tech). These traps were located in oak, spruce and pine forests and were baited with lures (100% ethanol, plus α-pinene) (Inward 2019). Propylene glycol (65%) was used as the preservation fluid at the bottom of the traps. Specimens were morphologically identified and selected at random to obtain the same number of specimens per beetle species and forest type.

Table 1. Overview of the beetle species and number of the specimens per forest type included in the study along with information related to their origin and main fungal structures.

Forest type	Beetle species	Status	Adapted structures	Feeding mode
Spruce, oak	*Xylosandrus germanus*	Introduced	Mesonotal mycangia	Xylomycetophagous
	Xyleborinus saxesenii	Native	Elytral mycangia	Xylomycetophagous
Pine, spruce	*Gnathotrichus materiarius*	Introduced	Tubular opening near precoxae	Xylomycetophagous
	Tomicus piniperda	Native	No known mycangia	Xylophagous

In the laboratory, the specimens were rinsed with pure water to remove loosely adhering fungal tissue, and thoroughly macerated individually to ensure that all fungi associated with the specimens were released. DNA was extracted using the Qiagen DNeasy Blood and Tissue spin column extraction kit (Qiagen, Valencia, CA, USA). Individual DNA extracts were first tested for correct species identification using the COI barcode marker, which was amplified for a 418 bp fragment and

sequenced on Illumina HiSeq following methods of Arribas (2016). In all cases the most abundant read, as determined with the NAPselect script (Creedy et al. 2019), had an exact match to existing reference sequences of the respective species, confirming the morphological identification.

The DNA extracts were then used for fungal metabarcoding of ITS with primers ITS86F (5′-GTGAATCATCGAATCTTTGAA-3′) / ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) (Op De Beeck et al. 2014) and LSU using primers LR0R (5′-ACCCGCTGAACTTAAGC-3′) (Vilgalys and Hester 1990) / JH-LSU-369rc (5′-CTTCCCTTTCAACAATTTCAC-3′) (Li et al. 2016) targeting the D1-D2 region at the 5' end of the LSU gene immediately downstream of the ITS2 region. Both markers were amplified from each beetle DNA extraction in separate reactions. Unique six-nucleotide indices added to each primer pair were used to distinguish the libraries. PCRs were pooled from three replicates conducted under slightly different annealing temperatures (54°C, 55°C and 56°C) to accommodate differences in optimal amplification conditions of the fungal species (Schmidt et al. 2013) and blank PCR reactions were used as negative control. Successful PCR amplicons were purified using the AMPure XP magnetic beads (Beckman Coulter) and amplicons were indexed using a secondary PCR with Nextera XT DNA Library Preparation Kit (Illumina Inc.) and sequenced on an Illumina® HiSeq® 2500 platform to generate 2 x 300 bp paired-end reads.

Bioinformatics

Raw reads were demultiplexed, primer trimmed and singleton reads removed with Cutadapt (Martin 2011). Read quality was evaluated using FastQC (Andrews et al. 2010). The raw reads generated for these analyses are available as Bio-Project PRJNA727174 (Sequence Read Archives (SRA)) in the BioSample Submission Portal (Barrett et al. 2012).

Forward and reverse reads were merged and quality filtered (Phred score ≥30) using PEAR v. 0.9.8 (Zhang et al. 2014). After merging, the average read length was 252 bp for ITS and 357 bp for LSU. Subsequent steps were carried out using VSEARCH v. 2.15.0 (Rognes et al. 2016) . A further quality test was effectuated using, with the --fastx_filter command and -fastq_maxee 1.0. After dereplication (--derep_fulllength), assemblies were denoised (--cluster_unoise --minsize 4 --unoise_alpha 2) and length filtered for a range of 100 to 500 bp (--fastx_filter) and all singletons removed. Chimera filtering was performed (--uchime3_denovo) and reads were then clustered into Operational Taxonomic Units (OTUs) at various similarity thresholds (97%, 98%, 99%) using VSEARCH. The average length of the OTUs was 270 bp for ITS and 347 bp for LSU (Supplementary Figure S1). Reads were then mapped to the OTU clusters, outputting an OTU table of read abundances suitable for the ecological analysis of the samples.

OTU identification and classification

Fungal OTUs were classified following three widely used methods for species identification. The Ribosomal Database Project (RDP) Bayesian Classifier (Wang et al. 2007) was used for fungal identification employing the Warcup fungal ITS (v2, release March 2018) and UNITE (version February 2020) training sets (Deshpande et al. 2016, Edgar 2018). In addition, OTUs were processed through the PROTAXFUNGI pipeline (Abarenkov et al. 2018), implemented in the PlutoF (Abarenkov et al. 2010) platform and based on the UNITE fungal database (February 2020). PROTAXFUNGI hierarchically assigns the OTU identities from the root node of the taxonomy through to the species (Nilsson et al. 2019). It has not been implemented for LSU, and thus was applied to the ITS data only. A third classifier, IDTAXA, employs machine learning to reduce over-classification errors to obtain a higher accuracy (Murali et al. 2018). Taxonomic assignment was carried out separately on class, order, genus, and species level. The percentage accuracy of identification from these methods was averaged as a measure of confidence of each identification (average of three values for ITS and two for LSU data). A minimum threshold of 70% probability for at least one of the classifiers was set, below which the OTUs were considered as "unclassified", together with other sequences that were identified with high confidence against database entries labelled as "unclassified", "unidentified" or "*incertae sedis*". When identifications disagreed among the classifiers, the one with the highest probability was selected. Taxonomic composition of samples was presented as the number of OTUs assigned to a given taxonomic level in a barplot created with *ggplot2* in Rstudio (Wickham 2016).

In a more detailed study of OTU assignments in the ecologically important class Sordariomycetes, OTUs generated from clustering at 97%, 98%, and 99% similarity thresholds were classified using the Warcup training set and the RDP fungal data set for ITS2 and LSU, respectively (Deshpande et al. 2016). All OTUs with a confidence of assignment >80% to class Sordariomycetes were retained. Order-level assignments (the Sordariomycetes are split into 28 orders) with a confidence >50% were taxonomized, while all others were kept as "unclassified Sordariomycetes". To assess the effects of differing clustering thresholds on downstream taxonomic assignment, OTUs at each clustering threshold were also closed-reference clustered (i.e. only retaining sequences with hits in the reference set) against the composite LSU/ITS2 reference sequences used to construct the tree (Edgar 2010, Rognes et al. 2016).

Alignment and tree building in Sordariomycetes

Reference sequences for the class Sordariomycetes were downloaded from GenBank, consisting of species that were complete for at least 2/3 of the rRNA cluster composed of SSU, LSU, and ITS2

sequence (full list of accessions in Suppl. Table 1). 80% of species in this reference set had accessions covering all three regions. ITS2 reference sequences were processed through ITSx to eliminate redundancy in the concatenated alignment (Bengtsson-Palme et al. 2013).

The reference sequences and OTUs were aligned using MUSCLE (Edgar 2004) under default settings and the aligned matrices were concatenated. The concatenated three-gene alignment was then inspected in Mesquite and Geneious Prime (v 2020.0.4) and problematic accession sequences were removed. This alignment is available on TreeBase (www.treebase.org accession number S28904). The alignment was then partitioned for each marker region, and the best model for each partition was selected according to BIC values. Model testing, tree building, and ultrafast bootstrap approximation (n = 1000) were performed in IQ-Tree2 (Chernomor et al. 2016).

Phylogenetic diversity metrics

Phylogenetic distribution of ITS and LSU copies was assessed by metrics of clustering and overdispersion originally developed for community ecology (Webb et al. 2008). In the ideal case of capturing the same species with both markers, copies of ITS2 and LSU corresponding to the same species should be in close vicinity on the tree, i.e. the copies of each marker should be 'overdispersed' (more dispersed than a random phylogenetic structure). Deviations from this pattern can be assessed with the metrics calculating the Mean Pairwise Distances (MPD) and Mean Nearest Taxon Distances (MNTD) of each set (ITS2 and LSU). We report standardized values as the net relatedness index (NRI) and nearest taxon index (NTI) relative to null models of randomly distributed communities. Positive NRI and NTI scores indicate phylogenetic clustering, negative values indicate phylogenetic overdispersions, while random phylogenetic structure results in values not significantly different from zero (Webb et al. 2008). Calculations were performed with the *R* packages *picante*, *ape*, and *phylomeasures* (Webb et al. 2008, Tsirogiannis and Sandel 2016, Paradis and Schliep 2019).

Assessment of species richness and community composition

Community ecological analyses were carried out on samples rarefied to 1000 reads, which was sufficient for generating largely complete OTU sets as judged by species accumulation curves (Supplementary Figure 2), thus retaining 73 samples from the four species. Species accumulation curves were built with the *specaccum* function of the *vegan* package (Oksanen et al. 2013). An OTU table and species classification was generated for fungal communities separately from ITS2 and LSU sequencing, after singletons and doubletons reads were removed. Fungal OTU richness among samples was assessed with a Generalised Linear Model (GLM), built using the *lme4* package (Bates et al.

2015) of *R*, with fungal OTU richness as a response variable and beetle species and forest type as dependent variables. The Negative Binomial model was chosen, as it is suitable for over dispersed data. A post hoc pairwise comparison (Tukey HSD test at the 95% significance level) was carried out to compare the means among the different factors.

Sorensen and Jaccard indexes were used to calculate beta-diversity between sample pairs based on presence-absence data (richness), and the Bray-Curtis index was used for turnover based on the relative abundance of each OTU. The variation was visualised using multidimensional scaling MDS (*betapart* *R* package; (Baselga and Orme 2012), and the significance of the distribution was assessed using permutational multivariate analysis of variance (PERMANOVA) under a model for forest type, beetle species and their interaction, calculated with the *adonis* function of the *vegan* package (permutations=999) (Oksanen et al. 2013). A permutational multivariate analysis of dispersion (PERMUTEST) was performed to evaluate the assumption of homogeneity of variance in the data by using the *betadisper* function (Anderson and Walsh 2013). The non-euclidean distances between the group centroids were also calculated to assess the group dispersion.

To evaluate the stringency of association of fungal OTUs with tree species and beetle hosts for each assembly, a multilevel pattern analysis was carried out by calculating Pearson's phi coefficient of association ("p.g") (Chytrý et al. 2002) between sample pairs, correcting this index to account for the differences in specimen numbers among the compared groups (function *multipatt* of the *indicspecies* *R* package; (De Cáceres et al. 2011). The identified OTUs for which the association values were significant were displayed as a heatmap (*aheatmap* function, NMF *R* package (Gaujoux and Seoighe 2010).

Results

Sequencing of 80 libraries produced 2,436,075 quality-filtered and merged reads for ITS2 and 1,742,119 reads for LSU, which resulted in 1157 OTUs from ITS2 and 548 OTUs from LSU after bioinformatics filtering and clustering (1546 and 632 OTUs if singleton and doubleton reads were retained and without applying rarefaction on each library). Identifications of OTUs at ≥70% confidence level obtained with IDtaxa, PROTAXfungi and RDP were higher for ITS2 than for LSU at all hierarchical levels from class to order, family, genus and species level (Fig. 1). Yet, OTUs identified by one or multiple identifiers never exceeded 61.5% for ITS and 41.5% for LSU of the total OTUs. Identifications dropped consistently from class to species level, and with each hierarchical level an

increasing proportion of identifications was provided by a single classifier only, indicating the growing uncertainty of taxonomic assignments. A classification at species-level was generally not possible for LSU because of the limitations of the databases that generally provide a taxonomy string to genus level only, and thus nearly 100% of the OTUs remained unidentified at this level. Nearly 50% of sequences were identified to species level with ITS but in virtually all cases only a single classifier produced these results (Fig. 1).

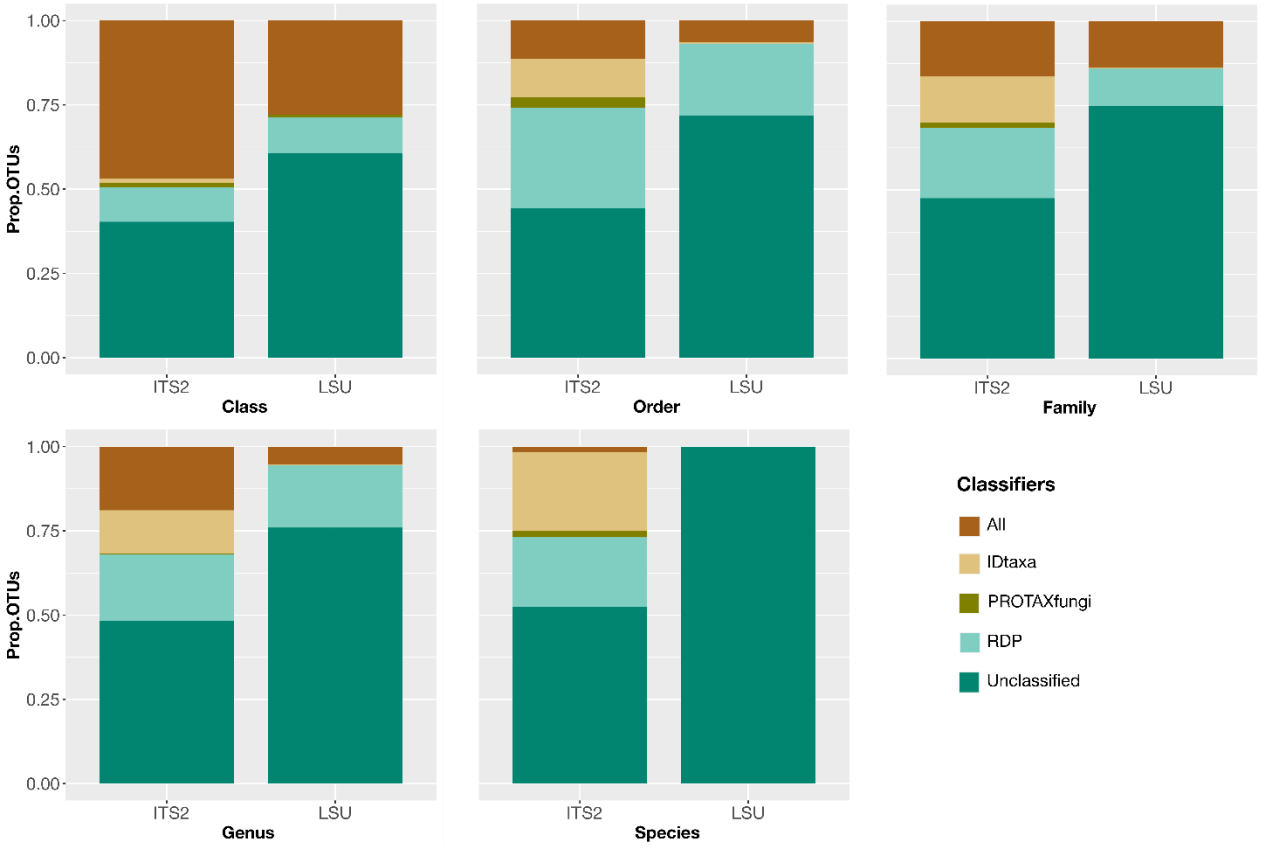

Figure 1. The proportion of fungi classified with IDtaxa, RDP and PROTAXfungi from class to species level. "ALL" refers to the proportion of OTUs for which the three classifiers agreed in their classification.

The 73 beetle specimens remaining after rarefaction harboured a total of 1180 OTUs for ITS and 553 OTUs for LSU. These OTUs could be assigned to 24 classes, 66 orders, 129 families and 369 genera. Identification at class level revealed the presence of 23 classes for ITS2 and 17 classes for LSU. The dominant classes were Dothideomycetes for ITS2 and Sordariomycetes for LSU (Fig. 2A, B, Table S1). ITS2 produced twice as many identified OTUs, and in the classes Leotiomycetes and Tremellomycetes more than five times as many, compared to LSU, due to the greater total number of OTUs and the higher proportion of them being fully identified. ITS2 also detected seven fungal classes not retrieved with LSU (Archaeorhizomycetes, Chytridiomycetes, Mucoromycetes, Orbiliomycetes,

Spizellomycetes, Tritirachiomycetes and Ustilaginomycetes), while LSU recovered only one class not obtained with the ITS2 (Atractiellomycetes). Only for the Sordariomycetes and Agaricomycetes the proportion of OTUs detected with LSU was higher than with the ITS2 marker.

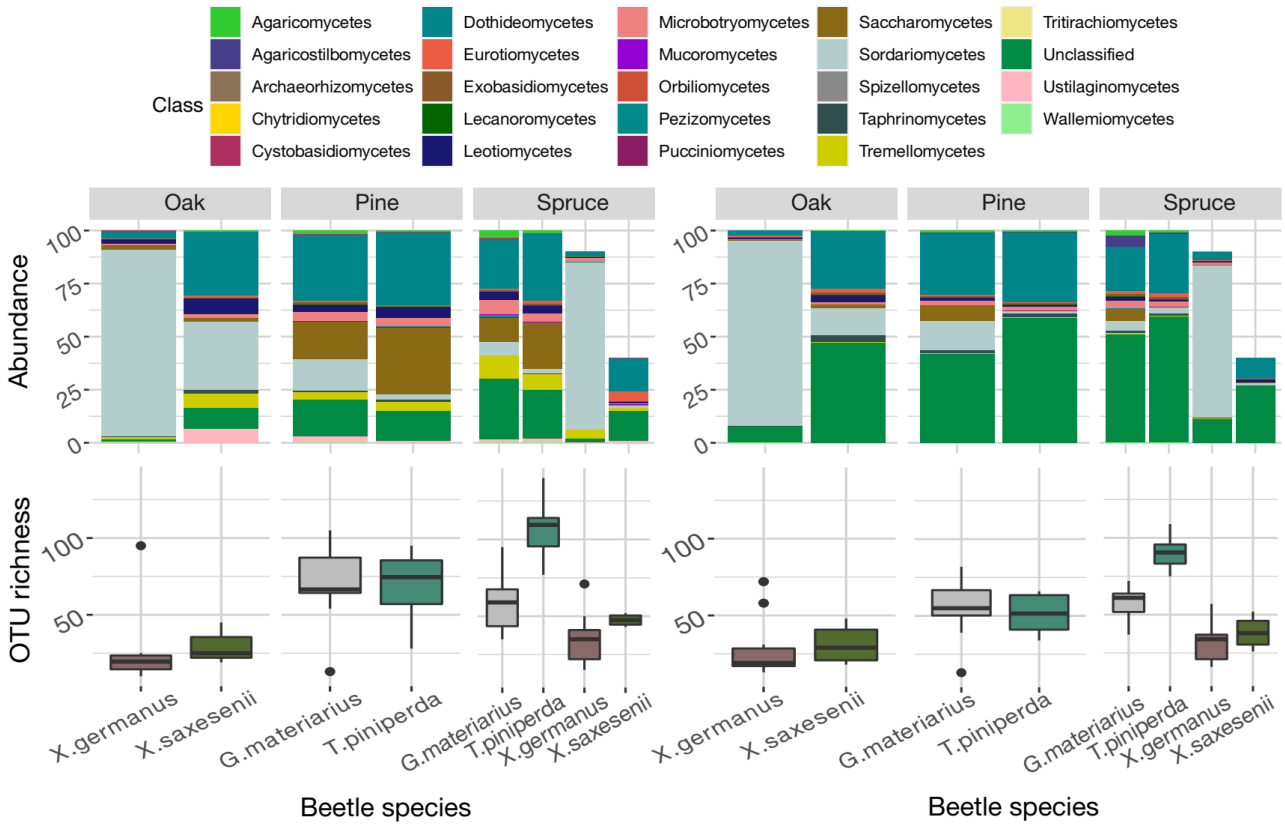

Figure 2. Proportion of OTUs identified as members of a fungal Class determined by the ITS (right panel, up) or LSU (left panel, up) regions. For the spruce forest, only nine and four *X. germanus* and *X. saxesenii* specimens, were retained after rarefaction. Number of fungal OTUs per beetle specimen, separate for each beetle species and forest type, for ITS (right panel, down) and LSU (left panel, down).

Effects of the amplified sub-region on the ecological analysis

Fungal communities obtained with either marker were compared with regard to total richness and differentiation across beetle species and forest type. For both markers, species accumulation curves displayed a similar shape, despite the roughly twice higher OTU number in ITS2, with a slow increase and not reaching a plateau, although LSU generally showed a more pronounced 'shoulder' indicating a fraction of OTUs that is encountered commonly in multiple samples. Across the different forest types, species accumulation in oak forest clearly lagged behind pine and spruce forests in both markers (Supplementary Fig. S2).

Richness in a single-beetle extract ranged from 9 to 140 fungal OTUs (mean of 56) in ITS2 and from 11 to 109 fungal OTUs (mean 48) in LSU (Fig. 2C, D). Despite some scatter among individual beetles, the number of OTUs per sample differed in a characteristic way between beetle species and forest types, and these differences were closely correlated in ITS2 and LSU, indicating that both markers detected a similar set of fungal species. This correlation was also evident at specimen level in the two outliers in each of the libraries corresponding to the same beetle individual. The variation in species richness explained by forest type and beetle species was broadly similar in ITS2 and LSU derived fungal communities (Table 2), but the LSU data attributed a greater proportion of the variation to the forest type alone (27.47% compared to 18.75% from ITS2), rather than to the interaction of forest type and beetle species.

Table 2. Overview of the GLM analysis results, showing the significant *p* values, the percentage of explained variance and the F parameter for every factor included in the model.

Factor	**Explained variance**		$F_{x,y}$		***p***	
	ITS2	**LSU**	**ITS2**	**LSU**	**ITS2**	**LSU**
Beetle + forest type	8.46%	7.58%	$F_{6,65} = 2.809$	$F_{6,65} = 3.521$	< .1 *	< .05 *
Beetle	27.07%	18.25%	$F_{3,69} = 30.265$	$F_{3,69} = 29.888$	< .001 ***	< .001 ***
Forest type	18.75%	27.47%	$F_{2,67} = 6.772$	$F_{2,67} = 5.236$	< .01 **	< .01 **
Unexplained	45.72%	46.68 %				

Both ITS2 and LSU explained around 7% of the variation to be due to beetle species. MDS plots on the OTU composition revealed a very similar pattern of community separation of the three forest types in ITS2 and LSU (Fig. 3).

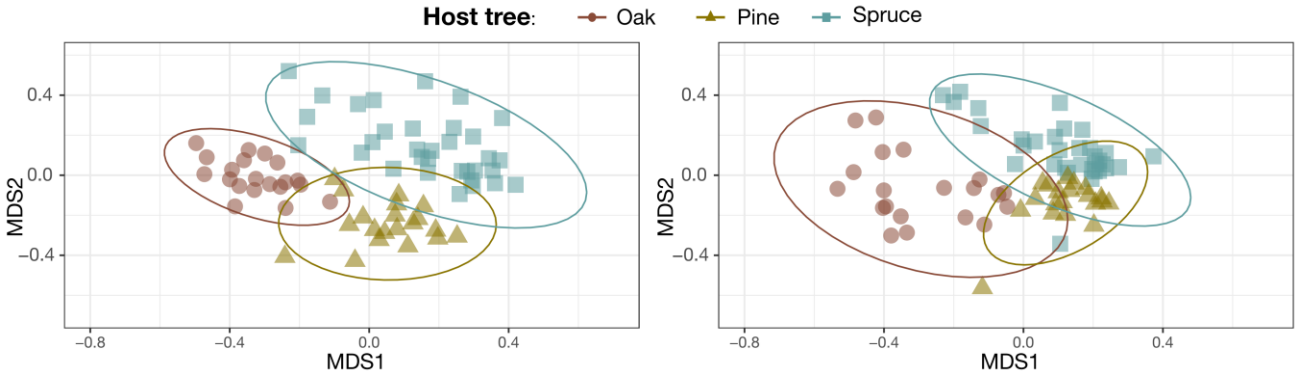

Figure 3. MDS ordination plot of all specimens sampled with ITS (right panel) or LSU (left panel) data, based on the fungal community composition of the individual beetles. Colours represent beetle species and shapes represent forest types. Stress for this graph fell within acceptable ranges (<0.2).

Tests of associations of OTUs to particular beetle species and forest types with the *indval* function revealed a large number of significantly positive and negative associations (Fig. 4). Despite representing fewer OTUs in total, the number of OTUs with significant levels of association was slightly higher in LSU compared to ITS2 (60 versus 50 OTUs). Positive associations of OTUs were prevalent in pine and spruce forests, but much less in oak. Positive associations were particularly strong with *T. piniperda*, and to some extent *G. materiarius*, whereas positive associations with *Xs. germanus* and *Xb. saxesenii* were limited to a small number of oak associated OTUs. Most other associations in these species were negative; e.g. the pine/spruce associated OTUs were absent, despite the fact that both beetle species were also sampled from spruce. General patterns of OTU associations and non-associations were similar for the two xyleborine species, and they were quite similar with those in Oak. In contrast, association patterns in *T. piniperda* and *G. materiarius* were similar to Pine and Spruce (Fig. 4). The similarity in these association patterns differed only slightly if assessed with the ITS2 and LSU-based OTUs (Fig. 4), despite the fact that the OTUs themselves could not be linked up between both markers, as they mostly were not identified or the identifications did not overlap between the two marker sets.

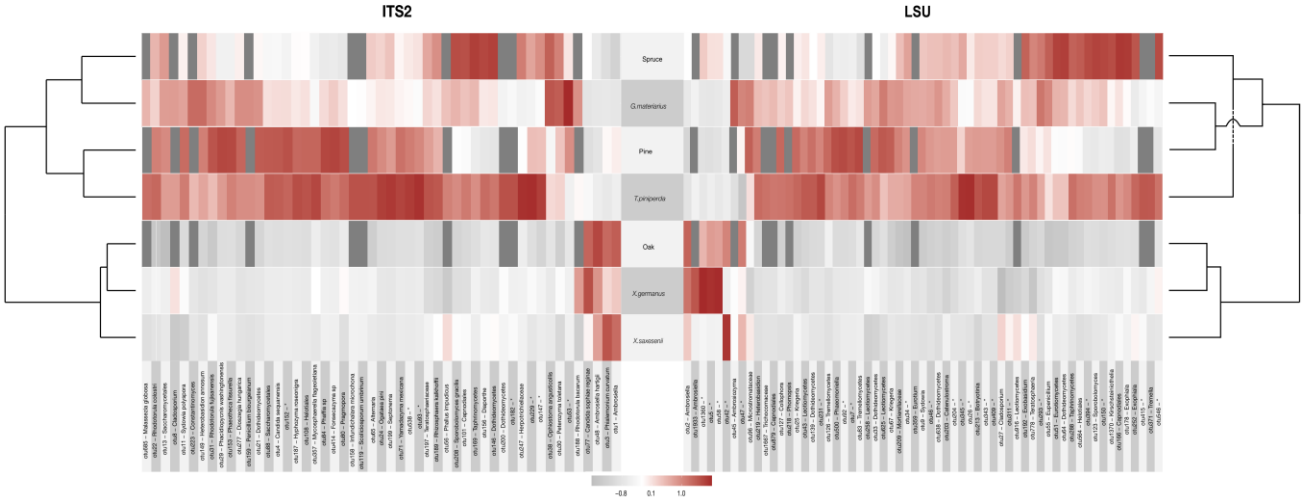

Figure 4. Heatmap using Pearson's correlation coefficient between the OTUs generated from the ITS and LSU metabarcodes and the analysed beetle species and Forest types. Squares indicate the strength of association between OTU and beetle/forest (red indicates relative high correlation and grey indicates relative low correlation). Fungal OTUs (on the x-axis) were classified to the genus or species level where possible; they are shown in a random order and cannot be linked taxonomically between both markers.

OTU identifications across markers

Because it remained unclear if the observed similarities in community composition were based on the same set of OTUs, a phylogenetic approach was used to associate ITS2-based and LSU-based OTUs with each other, focusing on the class Sordariomycetes as a model system. OTUs were clustered at minimum similarity thresholds of 97, 98 and 99%, which resulted in between 120-150 OTUs for ITS and 80-120 OTUs for LSU classified as Sordariomycetes using the RDP classifier at confidence >80% (Table 3). OTU sequences from either marker were used for phylogenetic analysis together with publicly available full-length sequences covering all or most of the rRNA cluster, including the LSU and ITS2 regions, besides the SSU gene present in most accessions. These sequences served as a scaffold for the major orders of Sordariomycetes (full list of accessions in Suppl. Table S1), to which the metabarcode data were added, using clustering thresholds of 99% similarity for LSU and 98% for ITS2. These conditions were used because they generated a similar number of OTUs for each marker (Table 3), and thus potentially represent a similar set of species.

Table 3. Net relatedness index (NRI), nearest taxon index (NTI), and number of OTUs recovered for LSU and ITS2 across all clustering thresholds. "Mixed" refers to a clustering threshold of 99% for LSU and 98% for ITS2. Reference sequences were included when building the trees used, though pruned (leaving only OTUs in the tree) for the above calculations.

	ITS2				**LSU**			
	97%	**98%**	**99%**	**Mixed**	**0,97**	**0,98**	**0,99**	**Mixed**
NRI	-0,111	-0,805	1,497	-0,122	1,328	0,697	-2,55	-0,653
NTI	-1,212	-3,81	-2,386	-2,277	1,367	0,882	-0,183	-1,343
OTU count	138	144	158	144	80	102	150	150

Maximum-likelihood trees for the three-gene reference alignment and OTUs from metabarcoding resolved basal relationships similar to those found in the literature (Hongsanan et al. 2017; Zhang et al. 2006) (Suppl. Fig. S3). OTUs obtained from ITS and LSU were widely distributed across this tree, and their placement generally matched the order-level assignment obtained with the RDP classifier, which was dominated by the orders Hypocreales, Diaporthales, Ophiostomatales and Xylariales (Fig. 5) (Suppl. Fig. S4). Across most orders in the trees, ITS2 and LSU sequences were recovered in close proximity and the distribution of both types of sequences across orders was similar, showing that overall community diversity at the order level could equally be inferred using either ITS2 or LSU (Suppl. Fig. S2).

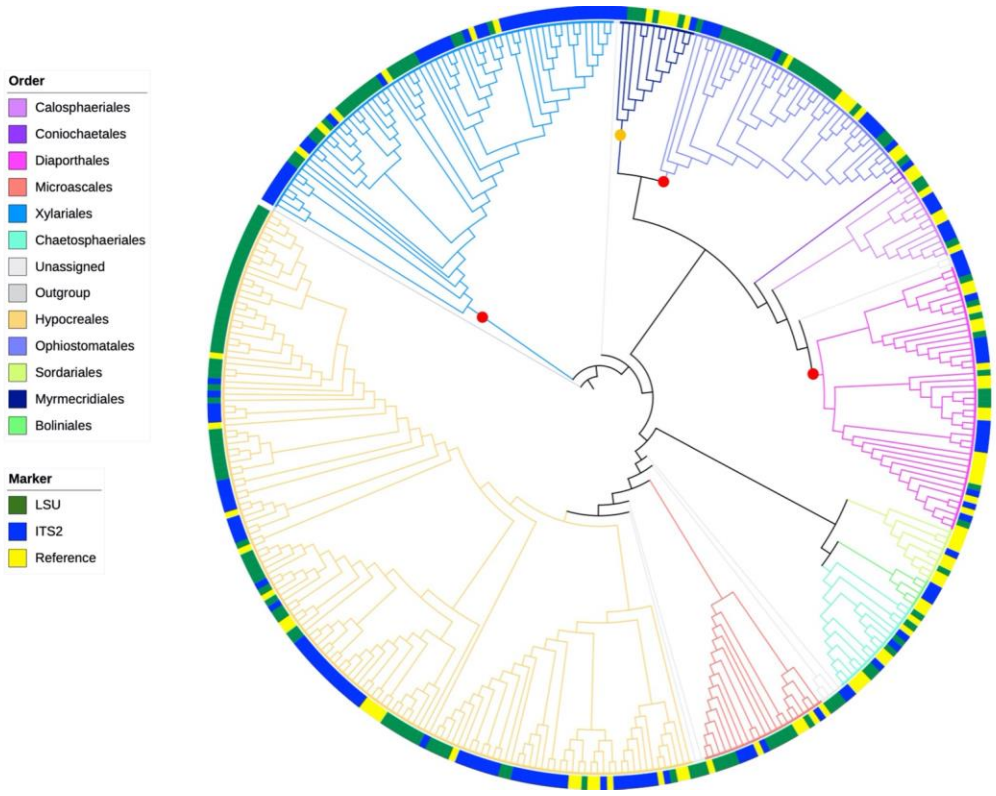

Figure 5. Phylogenetic tree of Sordariomycetes showing the orders represented by the colour of branches, and the distribution of ITS (blue) and LSU (green) relative to the reference set (yellow) on the outer circle. Note the limited presence of ITS sequences in the Ophiostomatales (in top right quadrant). Red dots mark the three orders Xylariales, Ophiostomatales and Diaporthales shown in greater detail in Figure 6. The orange dot marks the Myrmecridiales whose presence in the sample was only detected with the tree-based method, but not with the classifiers.

Order-level subsets of trees for these orders showed the placement of ITS2 and LSU sequence fragments relative to the reference set (Fig. 6A-C). If both sequences are derived from the same genomic template in the metabarcoding amplification they were expected to be represented by one OTU representative sequence for each marker, and these sequences to fall in close proximity on the tree (taking the same phylogenetic position) (Fig. 6D). However, due to their non-overlapping sequence, they can only be placed relative to the nearest full-length reference sequence. We found 15 instances where one ITS and one LSU barcode were in close proximity together with a reference sequence, potentially representing the same species. In an additional six instances, one or both of the

barcodes formed a cluster on zero-length branches when matched to full length rRNA reference sequences, i.e. representing an exact match to an existing database entry.

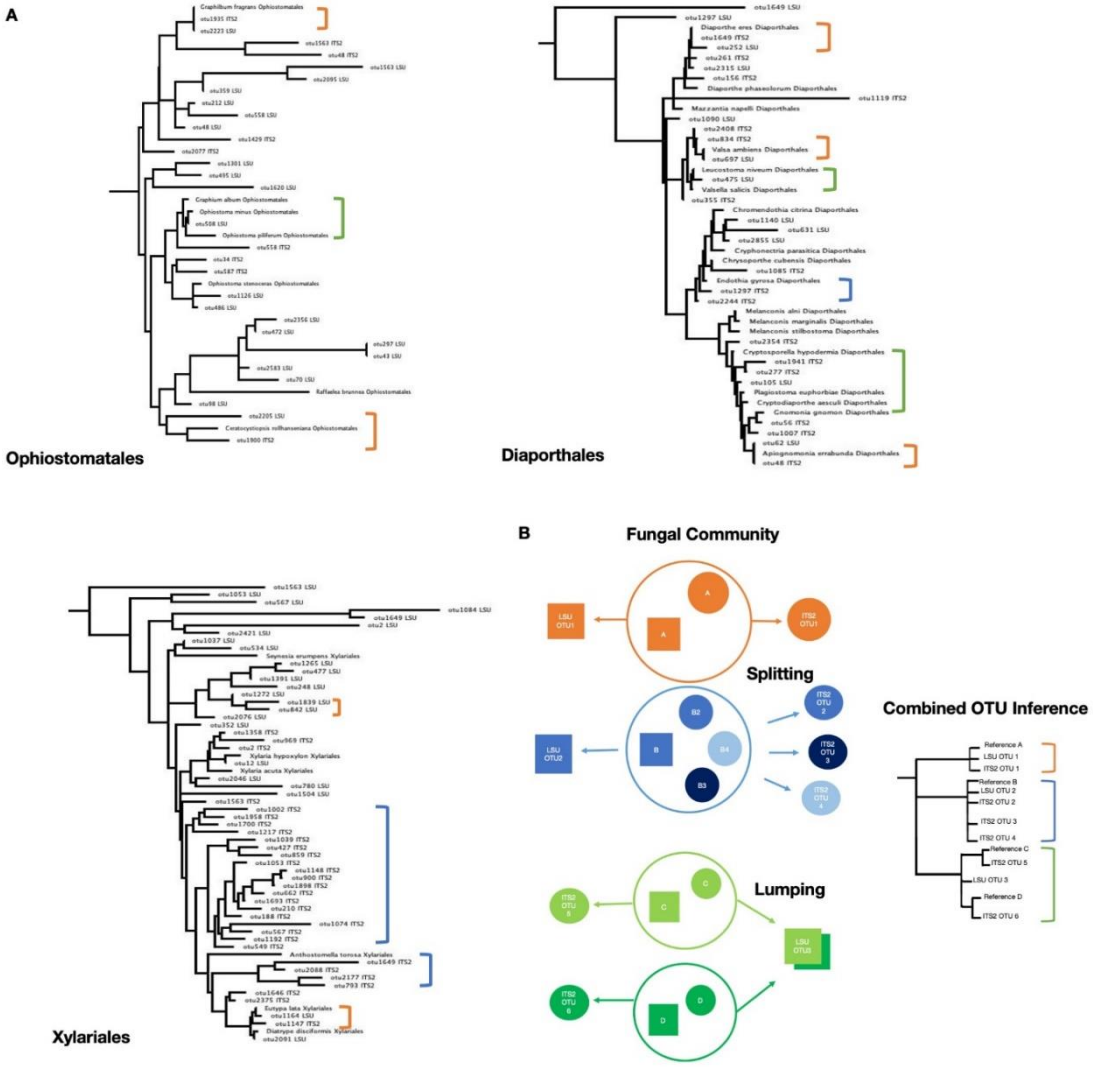

Figure 6. Order-level trees and splitting/lumping of OTUs at clustering. **A**. Order-level trees with mixed OTU clustering thresholds (99% LSU, 98% ITS2). Full tree in supplementary materials. *Leotia lubrica* was used as the outgroup (not pictured). Maximum likelihood trees were constructed in IQ-Tree2. Brackets indicate reference taxa linked to an ITS2 and/or LSU OTU, with colors indicating potential splitting/lumping (blue, splitting; green, lumping; orange, 1:1). **B**. Diagram illustrating the effects of splitting and lumping of an OTU in the fungal community on the tree inference. Four hypothetical species (A to D) in a community are treated under uniform clustering thresholds for ITS2 and LSU. This may result in deviation from the 1:1 ratio of OTUs expected if each species in the community is represented equally by both markers (species A). Threshold values may be too high resulting in splitting of species into multiples OTUs, which is likely to

affect the more variable ITS2 region (species B) or may be too low, resulting in lumping of multiple species into a single OTU, likely to affect the conservative LSU region (species C and D). Note that the interaction of non-overlapping ITS2 and LSU sequences in the combined community tree is only possible through their placement relative to the full-length reference sequences that include both markers. The combined OTU inference shows how the different clustering artefacts may present themselves on the tree.

Closed-reference clustering against the reference dataset to each order within Sordariomycetes by the RDP classifier revealed species-level matches for both ITS2 and LSU sequences (Fig. 7A). Notably, four species had matches to both markers, i.e. the same species were amplified. In addition, one ITS2 sequence produced a hit not reciprocated in LSU. Vice versa, LSU sequences produced hits to a minimum of eight additional species not seen in ITS, which was increased to 11 and 17 species under the higher 98 and 99% threshold values, respectively, as the trees became increasingly populated with the additional taxa from splitting of larger OTUs (Fig. 7B). Under these lower threshold values closely related sequences apparently were less affected by 'lumping', which obscured the true diversity in the sequencing mixture.

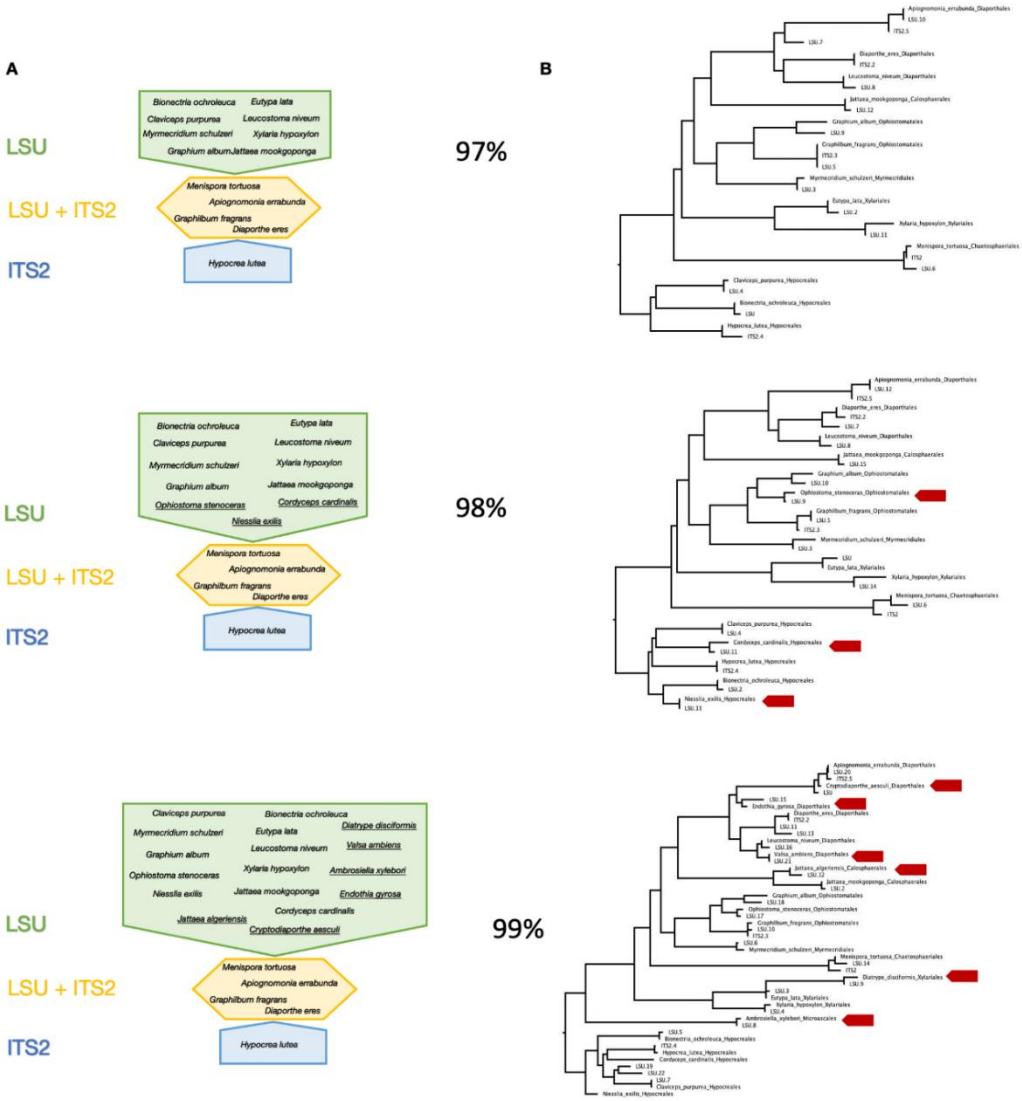

Figure 7. Closed reference clustering of OTUs and phylogenetic trees at different thresholds. **A**. Results from the closed reference clustering of OTUs at each clustering threshold against composite LSU/ITS2 reference sequences. LSU matches in green, ITS2 matches in blue, linked matches (for which both an ITS2 and LSU OTU were matched to a reference sequence of the same species) in yellow. Underlined taxa indicate new matches at each clustering threshold. **B**. Phylogenetic tree of LSU OTUs under increasingly stringent clustering thresholds, with arrows marking newly added taxa as threshold values are increased.

Where closely related reference sequences were missing, ITS2 and LSU sequences may be matched based on their phylogenetic proximity, but the ITS2 and OTU sequences obtained from a single genome may not appear as sister taxa because the gene sequences are non-overlapping and thus lack characters that could group them. We tested the degree to which ITS2 and LSU sequences interleave on the tree, by assessing phylogenetic clustering and dispersion with the NRI and NTI (Table 3). For

ITS2, most values were negative, indicating overdispersion relative to the LSU sequences, except for the 99% similarity value, which produced positive NRI (clustering) possibly from selective oversplitting of OTUs that is not matched in the less variable LSU sequences. For LSU there was a progression from positive (clustering) at 97% similarity to negative (indicating overdispersion) at 99% similarity, which coincided with a near doubling in the number of OTUs (against only a small increase in the ITS2 data) (Table 3). This indicated that OTUs newly formed by splitting were not clustered, unlike the ITS2-derived OTUs, but instead were widely distributed on the tree and interleaved with the ITS2 sequences. A 'mixed' threshold value of 98% for ITS2 and 99% for LSU presented slightly negative NRI/NTI values for both markers (Table 3).

The detailed observations were confirmed by the global classification of OTUs at order level, which showed an increase in the proportion of identified OTUs with increasing threshold value for LSU, but not ITS2 (Fig. 8). Both markers produced broadly similar proportions of the four dominant orders, Xyliariales, Ophiostomatales, Diaporthales and Hypocreales, but differed to various degrees in the assignment of the 'small' orders. It was also evident that OTU numbers in Ophiostomatales were comparatively lower in ITS2, as also suggested from the phylogenetic tree (Fig. 5). This may be explained by the fact that the ITS2 forward primer binding site in this group differs from the consensus (Suppl. Fig. S4)**.** The RDP classifications and tree-based assignments were largely in agreement regarding the relative proportion of order-level taxa, although generally the trees assigned a greater proportion of OTUs, reaching nearly 95%, and the result was more variable under the various similarity threshold values. This was particularly evident in ITS2, which saw hardly any change in the RDP classification under changing threshold values.

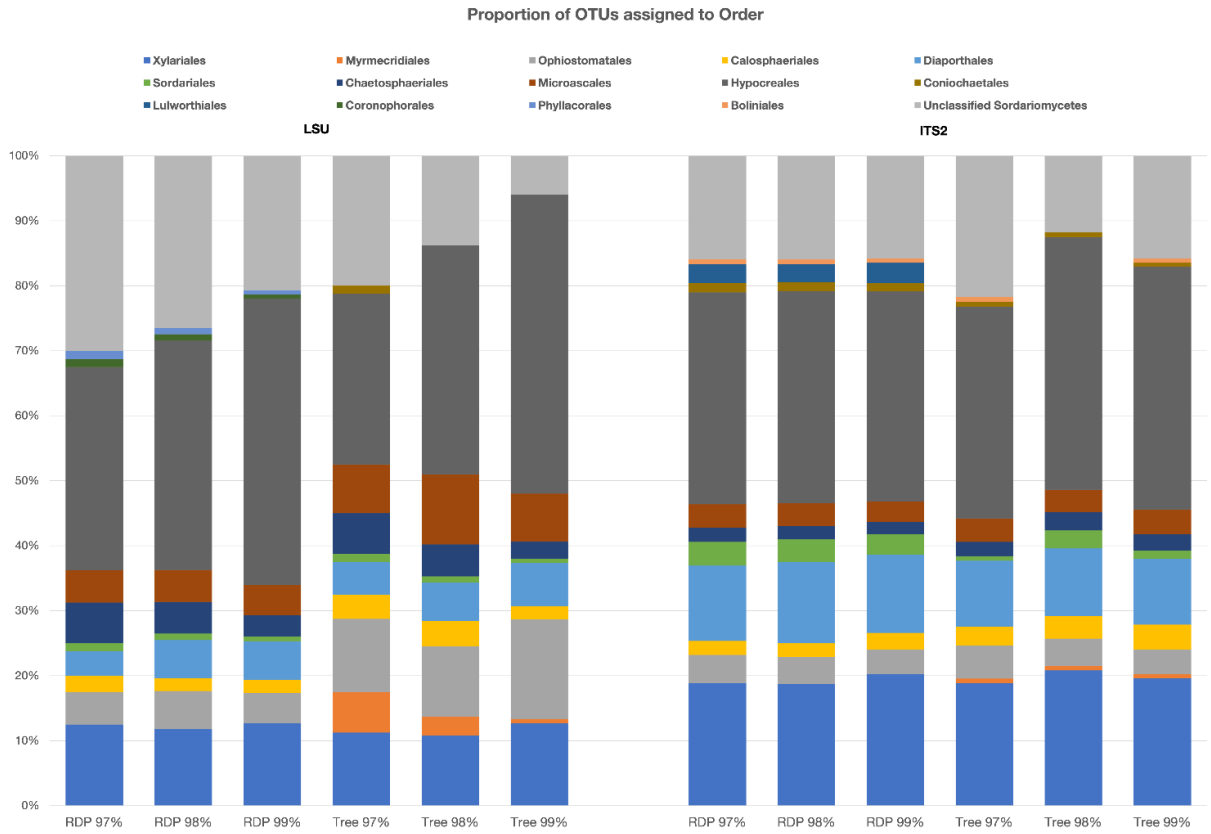

Figure 8. Classification of OTUs from metabarcoding with LSU (left panel) and ITS (right panel) markers based on the RDP classifier and the phylogenetic tree, under increasing threshold values.

Discussion

Metabarcoding has revolutionised the study of fungal communities, revealing the huge proportion of hitherto unobserved species, including the unexpectedly diverse communities associated with bark beetles (Tedersoo et al. 2019). However, these inferences are based on short sequences and lack the biological underpinning of conventional assessment of communities using fungal cultures. Independent corroboration of species limits is needed, and principally can be achieved by using multiple markers that confirm the same groupings independently (e.g. DeSalle et al. 2005). This was attempted here, in part motivated by the claims that the most widely used ITS fungal barcode marker may not be suitable for efficient amplification of key groups of fungi (Hulcr et al. 2020). The test of phylogenetic congruence is complicated because the amplicons come from complex mixtures of species, which does not allow to establish genetic linkage (phasing) across the two markers, despite the close proximity of the ITS2 and LSU in the genome. Instead, an indirect approach had to be used that identifies the amplicons relative to full-length reference sequences separately for ITS2 and LSU,

which was not straightforward given the highly incomplete and taxonomically incompatible databases available for both markers.

We addressed the equivalency of ITS and LSU for the characterisation of fungal communities, by (1) assessing the ecological associations at the higher taxonomic level for the entire fungal set, and (2) establishing explicit correspondences of species-level taxa in metabarcoding from both markers (for the class Sordariomycetes only). As we showed with both approaches, OTU identification is challenging and depends on the available reference databases, as well as the specific strategy for linking the metabarcode sequences into the taxonomic system. Taxonomic classifiers are now widely used and are becoming increasingly sophisticated. However, we note that placement is possible mostly to higher taxonomic levels, in line with comparisons of taxonomic classification software showing that the RDP classifier performed best in taxonomic classification to order (Richardson et al. 2017). Only a fairly small proportion of reads can be placed to genus or species level, in particular for the LSU marker (Fig. 1). Of those OTUs that could be placed, the exact taxonomic membership could be established only with one of the three classifiers. These difficulties in identifying species greatly compromised the comparison of community composition obtained with either marker, which was impossible to link with this approach, as virtually none of the species level entities encountered were labelled with the same Linnaean binomial (Fig. 3). Thus, we did not obtain a corroboration of community composition from using both markers via taxonomic identification of reads against their respective reference databases.

In contrast, simple counts of OTU numbers (a proxy of species richness) produced a good correlation between both markers in several key parameters describing the community composition. First, plots of the number of OTUs obtained from each treatment (beetle species, forest type) from the ITS2 and LSU data closely mirror each other (Fig. 1). Equally, the proportion of explained OTU diversity by beetle species, forest type and beetle x forest interactions was closely similar between both markers, even if the absolute number of OTUs was generally much lower in LSU (Table 2). For both markers, communities from different forest types occupy remarkably similar portions of the multivariate space (Fig. 4). In addition, the higher-level composition of fungal communities associated with each beetle species is very similar whether generated with ITS2 or LSU (Fig. 2), which also holds for the composition of classes within the order Sordariomycetes (Fig. 8). The broad patterns of individual OTU associations in the *indval* analysis also show similar affinities with the beetle species and tree type (Fig. 4), even if the correspondences of species identity between ITS2 and LSU datasets could not be determined. All of these findings point to a high level of corroboration between both markers and provide justification for the widely used approach of fungal community analysis based on higher

level classification and read abundances. The latter is particularly remarkable given the frequently raised concern about PCR bias in relative read abundance (Bálint et al. 2016, Krehenwinkel et al. 2017). Thus, even a single metabarcode marker can safely represent the broad ecological trends determining fungal communities.

Yet, the difficulty of linking these metabarcoding sequences amongst markers leaves great uncertainty about the biological relevance of the community data, which still may represent entirely different species within the major taxonomic groups recovered by either marker, as already suggested for the Ophiostomatales (Hulcr et al. 2020). Thus, ultimately the metabarcoding approach may fall short of linking any particular fungal species to a beetle, unlike the conventional approaches of culturing specific isolates that were used to reveal the symbioses of particular fungal and beetle species of direct relevance to the ecological processes. Phylogenetic analysis of individual sequences can improve the precision of placement, beyond the assignment to a broad taxonomic group, and using the phylogenetic position of particular ITS2 and LSU sequences relative to each other can potentially link the corresponding reads representing a given species from either marker, thus reconciling the distinct 'views' of the true community composition as revealed by each marker (Fig. 6D).

As shown by our extensive analysis of the Sordariomycetes, phylogenetic analyses provide clear insight into the taxonomic makeup of fungal communities and the most appropriate treatment of fungal metabarcodes for species delimitation. First, we found that the classifiers are broadly in agreement with the phylogenetic analysis. However, average RDP classifier confidence scores varied from order to order, with LSU assignments less confident overall, ranging from an average of 41% in Sordariomycetes, to values generally between 80-100% in the ITS sequences (see Suppl. Table S2). Some OTUs were not identified beyond the class level by the RDP classifier, despite clear placement in the tree. Notably, the order Myrmecridiales was missing entirely from the classifier results, despite the presence of several OTUs placed clearly within the order and OTUs matching *Myrmecridium schulzeri* found in the closed-reference clustering at all three thresholds (Fig. 7A, B). This outcome highlights the potential shortcomings of using classifiers that are dependent on reference databases and probabilistic assessments of short *k-mers*, which may be particularly problematic due to the limited sequence length of metabarcoding reads (Wang et al. 2007, Porras-Alfaro et al. 2014, Bacci et al. 2015, Xue et al. 2019).

Second, the phylogenetic analysis attempted to determine if both markers reveal the same species-level entities. Under ideal circumstances, each species is represented by exactly one sequence each of LSU and ITS, and these two sequences from both markers find themselves in the same position of the

tree. Thus, the 1:1 represented sequences of ITS and LSU should be interleaved on the tree, which predicts a uniform phylogenetic distribution of both markers, contrary to a scenario of overdispersion or clustering of either type (Fig. 6D). However, deviations from the uniform distribution are due to two factors; the non-overlapping sequence data that only can place ITS2 and LSU sequences relative to full-length reference sequences rather than to each other, and the problem of similarity thresholds that produce different numbers of potential species representatives by either lumping (too few species) or oversplitting (too many species) of OTUs in one or both of these loci. We here used the NRI/NTI framework to assess the uniformity of distributions of ITS and LSU sequences relative to each other on the phylogenetic tree.

Overall, the increase of the similarity threshold had a greater impact on the LSU than ITS2, almost doubling in numbers of recognised OTUs versus a small increase only (Table 3), and parity of OTUs in both markers was greatest at a 'mixed' threshold of 99% for LSU and 98% for ITS2. Slightly negative NRI/NTI values under these OTU thresholds (Table 3) approximate a pattern of overdispersed 'communities' of LSU and ITS2 sequences, as expected for the 1:1 correspondence of each marker, although the exact relationships were more complex. The 99% threshold for LSU is also supported by the greater matches in the closed-reference clustering (Fig. 7). The similarities in OTU counts and relative NRI/NTI values thus were considered the best estimate of the OTU diversity in each marker. Placement in proximity on the tree therefore would suggest that the respective ITS2 and LSU sequences in fact are derived from closely related fungal species, if not from the same genomic template. Frequently in our analysis this was corroborated by the fact that these closely related ITS2 and LSU sequences were obtained from the same specimen sample (not shown).

There are uncertainties associated with this inference, as it is well known that there is great variation in the rate of sequence evolution. Across many fungal classes, intraspecific ITS variability varies considerably, highlighting the challenges and inevitable shortcomings resulting from the selection of a uniform OTU clustering threshold (Nilsson et al. 2008). For example, while a 97% clustering threshold is generally accepted (Kõljalg et al. 2013, Tedersoo et al. 2015b), in one study intraspecific ITS variability in *Xylaria hypoxylon* was found to be 24.2%. Thus, the abundance of ITS2 OTUs in Xylariales relative to LSU OTUs (Fig. 6) may be the result of oversplitting of highly variable ITS2 OTUs (Nilsson et al. 2008). Yet, whether these OTUs correspond to biological species or some other hierarchical level, the approach generally is capable of linking close relatives across different markers. Reference sequences spanning both markers are required, but as shown for the Sordariomycetes, even an incomplete set can provide the scaffold for placing non-overlapping sequences, and in several instances the idealized placement of OTUs and reference sequences was found, in some cases across

all clustering thresholds (Fig. 5), while in many cases some uncertainties remain where reference sequences are distant. Matching of ITS2 and LSU sequences was even possible in the Ophistomatales, despite the deviation in the ITS2 primer binding site in this group (Suppl. Fig. S4), as the bias against the amplification presumably is overcome by permissive PCR conditions, and similar effects can be expected in other groups where such variation may exist, although we did not find any evidence in the taxonomically broad reference sequences used here.

Conclusion

Two streams of evidence suggest the equivalency of the two most widely used metabarcoding markers, ITS2 and LSU, in assessing fungal communities associated with insects: community-level diversity metrics showed great consistency when inferred with either marker, and the taxonomic and phylogenetic distribution of taxa recovered were highly similar. Metabarcoding of these communities only began recently, revealing a huge diversity of mostly anonymous ("dark") fungal isolates that have been used to draw conclusions on biodiversity patterns, but the taxonomic assignment and estimates of species richness were lacking independent corroboration. The use of two independent metabarcode markers changes this situation dramatically, as each inference is independent, and although the markers are genetically linked, i.e. likely represent the same evolutionary history, the experimental approach is treating these samples as separate inferences of the fungal communities. This provides strong support for the metabarcoding approach of fungal diversity directly from the insect specimens, which can reveal critical information about the distribution, spread and epidemic conditions of fungal tree pathogens by sampling the insect specimens.

Funding

The financial support for this project was provided by the John Spedan Lewis Foundation (JSLF).

Competing interests

The authors have declared that no competing interests exist.

Acknowledgements

The authors have no support to report.

Supplementary material

Supplementary Table S1. Class level identification of OTUs showing the number of OTUs produced with ITS2 and LSU and the proportion of the total OTU set on the rarefied data. OTUs classified at species level but not correctly classified at class level were considered as "Misclassified".

Class	n.OTUs.ITS2	Prop.OTUs.ITS2	n.OTUs.LSU	Prop.OTUs.LSU	ITS2:LSU_ratio
Agaricomycetes	48	4.07	34	6.15	1.41
Agaricostilbomycetes	2	0.17	1	0.18	2.00
Archaeorhizomycetes	1	0.08	0	0	0
Chytridiomycetes	1	0.08	0	0	0
Cystobasidiomycetes	6	0.51	2	0.36	3.00
Dothideomycetes	118	10	31	5.61	3.81
Eurotiomycetes	48	4.07	12	2.17	4.00
Exobasidiomycetes	16	1.36	3	0.54	5.33
Lecanoromycetes	16	1.36	4	0.72	4.00
Leotiomycetes	90	7.63	11	1.99	8.18
Microbotryomycetes	19	1.61	2	0.36	9.50
Mucoromycetes	17	1.44	0	0	0
Orbiliomycetes	2	0.17	0	0	0
Pezizomycetes	16	1.36	1	0.18	16.00
Pucciniomycetes	3	0.25	3	0.54	1.00
Saccharomycetes	37	3.14	6	1.08	6.17
Sordariomycetes	79	6.69	45	8.14	1.76
Spizellomycetes	1	0.08	0	0	0
Taphrinomycetes	9	0.76	1	0.18	9.00
Tremellomycetes	72	6.1	7	1.27	10.29
Tritirachiomycetes	1	0.08	0	0	0
Ustilaginomycetes	8	0.68	0	0	0
Wallemiomycetes	4	0.34	1	0.18	4.00
Atractiellomycetes	0	0	1	0.18	0
Misclassified	112	9.49	62	11.21	2.20
Unclassified	454	38.47	326	58.95	1.39
Total:	1180	1	553	1	
	Classified:	61,53		41,05	

Supplementary Table S2. Number of OTUs assigned to each order based on RDP Bayesian classifier. Average confidence scores were calculated over all order-level assignments, though only classifications with >50% confidence were taxonomized, all other were kept as "unclassified Sordariomycetes".

	LSU				ITS2			
Order	**0,97**	**0,98**	**0,99**	**Average confidence**	**0,97**	**0,98**	**0,99**	**Average confidence**
Xylariales	10	12	19	0,79	26	27	32	0,9
Myrmecridiales	0	0	0	n/a	0	0	0	n/a
Ophiostomatales	4	6	7	0,61	6	6	6	0,93
Calosphaeriales	2	2	3	0,98	3	3	4	0,99
Diaporthales	3	6	9	0,74	16	18	19	0,96
Sordariales	1	1	1	0,41	5	5	5	0,8
Chaetosphaeriales	5	5	5	0,98	3	3	3	0,99
Microascales	4	5	7	0,79	5	5	5	0,99
Hypocreales	25	36	66	0,8	45	47	51	0,92
Coniochaetales	0	0	0	n/a	2	2	2	1
Lulworthiales	0	0	0	n/a	4	4	5	0,94
Coronophorales	1	1	1	0,97	0	0	0	n/a
Phyllacorales	1	1	1	0,93	0	0	0	n/a
Boliniales	0	0	0	n/a	1	1	1	0,79
Unclassified Sordariomycetes	24	27	31	n/a	22	23	25	n/a

References

Abarenkov K, Somervuo P, Nilsson RH, Kirk PM, Huotari T, Abrego N, Ovaskainen O (2018) PROTAX - fungi: a web-based tool for probabilistic taxonomic placement of fungal internal transcribed spacer sequences. New Phytologist 220: 517–525. https://doi.org/10.1111/nph.15301

Abarenkov K, Tedersoo L, Nilsson RH, Vellak K, Saar I, Veldre V, Parmasto E, Prous M, Aan A, Ots M (2010) PlutoF—a web based workbench for ecological and taxonomic research, with an online implementation for fungal ITS sequences. Evolutionary Bioinformatics 6: EBO-S6271. https://doi.org/10.4137/EBO.S6271

Agerbo Rasmussen J, Nielsen M, Mak SS, Döring J, Klincke F, Gopalakrishnan S, Dunn RR, Kauer R, Gilbert MTP (2020) eDNA-based biomonitoring at an experimental German vineyard to characterize how management regimes shape ecosystem diversity. Environmental DNA. https://doi.org/10.1002/edn3.131

Anderson MJ, Walsh DCI (2013) PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? Ecological Monographs 83: 557–574. https://doi.org/10.1890/12-2010.1

Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S (2010) FastQC. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arribas P, Andújar C, Hopkins K, Shepherd M, Vogler AP (2016) Metabarcoding and mitochondrial metagenomics of endogean arthropods to unveil the mesofauna of the soil. Methods in Ecology and Evolution 7: 1071–1081. https://doi.org/10.1111/2041-210X.12557

Bacci G, Bani A, Bazzicalupo M, Ceccherini MT, Galardini M, Nannipieri P, Pietramellara G, Mengoni A (2015) Evaluation of the performances of ribosomal database project (RDP) classifier for taxonomic assignment of 16S rRNA metabarcoding sequences generated from Illumina-Solexa NGS. Journal of genomics 3: 36. https://doi.org/10.7150/jgen.9204

Bakker MG (2018) A fungal mock community control for amplicon sequencing experiments. Molecular ecology resources 18: 541–556. https://doi.org/10.1111/1755-0998.12760

Bálint M, Schmidt P-A, Sharma R, Thines M, Schmitt I (2014) An Illumina metabarcoding pipeline for fungi. Ecology and Evolution 4: 2642–2653. https://doi.org/10.1002/ece3.1107

Bálint M, Bahram M, Eren AM, Faust K, Fuhrman JA, Lindahl B, O'Hara RB, Öpik M, Sogin ML, Unterseher M, Tedersoo L (2016) Millions of reads, thousands of taxa: microbial community structure and associations analyzed via marker genes. van der Meer JR (Ed.). FEMS Microbiology Reviews 40: 686–700. https://doi.org/10.1093/femsre/fuw017

Barrett T, Clark K, Gevorgyan R, Gorelenkov V, Gribov E, Karsch-Mizrachi I, Kimelman M, Pruitt KD, Resenchuk S, Tatusova T (2012) BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. Nucleic acids research 40: D57–D63. https://doi.org/10.1093/nar/gkr1163

Baselga A, Orme CDL (2012) Betapart : an R package for the study of beta diversity: *Betapart package*. Methods in Ecology and Evolution 3: 808–812. https://doi.org/10.1111/j.2041-210X.2012.00224.x

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67. https://doi.org/10.18637/jss.v067.i01

Batra LR (1963) Ecology of Ambrosia Fungi and Their Dissemination by Beetles. Transactions of the Kansas Academy of Science (1903-) 66: 213. https://doi.org/10.2307/3626562

Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiology 10: 189. https://doi.org/10.1186/1471-2180-10-189

Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, de Sousa F, Amend AS, Jumpponen A, Unterseher M, Kristiansson E, Abarenkov K, Bertrand YJK, Sanli K, Eriksson KM, Vik U, Veldre V, Nilsson RH (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. Bunce M (Ed.). Methods in Ecology and Evolution: 914–919. https://doi.org/10.1111/2041-210X.12073

Chernomor O, Von Haeseler A, Minh BQ (2016) Terrace aware data structure for phylogenomic inference from supermatrices. Systematic biology 65: 997–1008. https://doi.org/10.1093/sysbio/syw037

Chytrý M, Tichý L, Holt J, Botta-Dukát Z (2002) Determination of diagnostic species with statistical fidelity measures. Journal of Vegetation Science 13: 79–90. https://doi.org/10.1111/j.1654-1103.2002.tb02025.x

Creedy TJ, Ng WS, Vogler AP (2019) Toward accurate species-level metabarcoding of arthropod communities from the tropical forest canopy. Ecology and evolution 9: 3105–3116. https://doi.org/10.1002/ece3.4839

De Cáceres M, Sol D, Lapiedra O, Legendre P (2011) A framework for estimating niche metrics using the resemblance between qualitative resources. Oikos 120: 1341–1350. https://doi.org/10.1111/j.1600-0706.2011.19679.x

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical transactions of the royal society B: Biological sciences 360: 1905–1916. https://doi.org/10.1098/rstb.2005.1722

Deshpande V, Wang Q, Greenfield P, Charleston M, Porras-Alfaro A, Kuske CR, Cole JR, Midgley DJ, Tran-Dinh N (2016) Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. Mycologia 108: 1–5. https://doi.org/10.3852/14-293

Dreaden TJ, Davis JM, de Beer ZW, Ploetz RC, Soltis PS, Wingfield MJ, Smith JA (2014) Phylogeny of ambrosia beetle symbionts in the genus Raffaelea. Fungal Biology 118: 970–978. https://doi.org/10.1016/j.funbio.2014.09.001

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32: 1792–1797. https://doi.org/10.1093/nar/gkh340

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

Edgar RC (2018) Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. PeerJ 6: e4652. https://doi.org/10.7717/peerj.4652

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

Frau A, Kenny JG, Lenzi L, Campbell BJ, Ijaz UZ, Duckworth CA, Burkitt MD, Hall N, Anson J, Darby AC, Probert CSJ (2019) DNA extraction and amplicon production strategies deeply influence the outcome of gut mycobiome studies. Scientific Reports 9: 9328. https://doi.org/10.1038/s41598-019-44974-x

Ganter PF (2006) Yeast and Invertebrate Associations. In: Péter G, Rosa C (Eds), Biodiversity and Ecophysiology of Yeasts. Springer, Berlin, Berlin, 303–370. https://doi.org/10.1007/3-540-30985-3_14

Gaujoux R, Seoighe C (2010) A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 11: 367. https://doi.org/10.1186/1471-2105-11-367

George PB, Creer S, Griffiths RI, Emmett BA, Robinson DA, Jones DL (2019) Primer and database choice affect fungal functional but not biological diversity findings in a national soil survey. Frontiers in Environmental Science 7: 173. https://doi.org/10.3389/fenvs.2019.00173

Harrington TC, Yun HY, Lu S-S, Goto H, Aghayeva DN, Fraedrich SW (2011) Isolations from the redbay ambrosia beetle, *Xyleborus glabratus* , confirm that the laurel wilt pathogen, *Raffaelea lauricola* , originated in Asia. Mycologia 103: 1028–1036. https://doi.org/10.3852/10-417

Hibbett D, Abarenkov K, Kõljalg U, Öpik M, Chai B, Cole J, Wang Q, Crous P, Robert V, Helgason T (2016) Sequence-based classification and identification of Fungi. Mycologia 108: 1049–1068. https://doi.org/10.3852/16-130

Hulcr J, Barnes I, De Beer ZW, Duong TA, Gazis R, Johnson AJ, Jusino MA, Kasson MT, Li Y, Lynch S, Mayers C, Musvuugwa T, Roets F, Seltmann KC, Six D, Vanderpool D, Villari C (2020) Bark beetle mycobiome: collaboratively defined research priorities on a widespread insect-fungus symbiosis. Symbiosis. https://doi.org/10.1007/s13199-020-00686-9

Hyde KD, Udayanga D, Manamgoda DS, Tedersoo L, Larsson E, Abarenkov K, Bertrand YJ, Oxelman B, Hartmann M, Kauserud H (2013) Incorporating molecular data in fungal systematics: a guide for aspiring researchers. arXiv preprint arXiv:1302.3244. https://doi.org/10.5943/CREAM/3/1/1

Inward DJG (2019) Three new species of ambrosia beetles established in Great Britain illustrate unresolved risks from imported wood. Journal of Pest Science. https://doi.org/10.1007/s10340-019-01137-1

Johnson AJ, McKenna DD, Jordal BH, Cognato AI, Smith SM, Lemmon AR, Lemmon EM, Hulcr J (2018) Phylogenomics clarifies repeated evolutionary origins of inbreeding and fungus farming in bark beetles (Curculionidae, Scolytinae). Molecular Phylogenetics and Evolution 127: 229–238. https://doi.org/10.1016/j.ympev.2018.05.028

Jusino MA, Banik MT, Palmer JM, Wray AK, Xiao L, Pelton E, Barber JR, Kawahara AY, Gratton C, Peery MZ (2019) An improved method for utilizing high-throughput amplicon sequencing to

determine the diets of insectivorous animals. Molecular Ecology Resources 19: 176–190. https://doi.org/10.1111/1755-0998.12951

Kandawatte Wedaralalage TC, Jayawardena RS, Hyde KD (2020) Hurdles in fungal taxonomy: Effectiveness of recent methods in discriminating taxa. Megataxa 1: 114–122. https://doi.org/10.11646/megataxa.1.2.2

Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AF, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM (2013) Towards a unified paradigm for sequence-based identification of fungi. https://doi.org/10.1111/mec.12481

Krehenwinkel H, Wolf M, Lim JY, Rominger AJ, Simison WB, Gillespie RG (2017) Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. Scientific reports 7: 1–12. https://doi.org/10.1038/s41598-017-17333-x

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

Li Y, Simmons DR, Bateman CC, Short DPG, Kasson MT, Rabaglia RJ, Hulcr J (2016) Correction: New Fungus-Insect Symbiosis: Culturing, Molecular, and Histological Methods Determine Saprophytic Polyporales Mutualists of Ambrosiodmus Ambrosia Beetles. PLOS ONE 11: e0147305. https://doi.org/10.1371/journal.pone.0147305

Lindgren B (1983) A multiple funnel trap for scolytid beetles (Coleoptera). The Canadian Entomologist, 115: 299–302.

Lücking R, Aime MC, Robbertse B, Miller AN, Ariyawansa HA, Aoki T, Cardinali G, Crous PW, Druzhinina IS, Geiser DM, Hawksworth DL, Hyde KD, Irinyi L, Jeewon R, Johnston PR, Kirk PM, Malosso E, May TW, Meyer W, Öpik M, Robert V, Stadler M, Thines M, Vu D, Yurkov AM, Zhang N, Schoch CL (2020) Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? IMA Fungus 11: 14. https://doi.org/10.1186/s43008-020-00033-z

Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G, Dentinger B, Padamsee M, Hibbett D, James TY, Baloch E (2004) Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. American journal of botany 91: 1446–1480. https://doi.org/10.3732/ajb.91.10.1446

Malacrinò A, Rassati D, Schena L, Mehzabin R, Battisti A, Palmeri V (2017) Fungal communities associated with bark and ambrosia beetles trapped at international harbours. Fungal Ecology 28: 44–52. https://doi.org/10.1016/j.funeco.2017.04.007

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17: 10. https://doi.org/10.14806/ej.17.1.200

Miller KE, Hopkins K, Inward DJG, Vogler AP (2016) Metabarcoding of fungal communities associated with bark beetles. Ecology and Evolution 6: 1590–1600. https://doi.org/10.1002/ece3.1925

Miller KE, Inward DJG, Gomez-Rodriguez C, Baselga A, Vogler AP (2019) Predicting the unpredictable: How host specific is the mycobiota of bark and ambrosia beetles? Fungal Ecology 42: 100854. https://doi.org/10.1016/j.funeco.2019.07.008

Murali A, Bhargava A, Wright ES (2018) IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome 6: 140. https://doi.org/10.1186/s40168-018-0521-5

Nilsson RH, Larsson K-H, Taylor AFS, Bengtsson-Palme J, Jeppesen TS, Schigel D, Kennedy P, Picard K, Glöckner FO, Tedersoo L, Saar I, Kõljalg U, Abarenkov K (2019) The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research 47: D259–D264. https://doi.org/10.1093/nar/gky1022

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H (2013) Package 'vegan.' Community ecology package, version 2: 1–295.

Op De Beeck M, Lievens B, Busschaert P, Declerck S, Vangronsveld J, Colpaert JV (2014) Comparison and Validation of Some ITS Primer Pairs Useful for Fungal Metabarcoding Studies. Neilan B (Ed.). PLoS ONE 9: e97629. https://doi.org/10.1371/journal.pone.0097629

Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology 18: 1403–1414. https://doi.org/10.1111/1462-2920.13023

Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35: 526–528. https://doi.org/10.1093/bioinformatics/bty633

Ploetz RC, Hulcr J, Wingfield MJ, de Beer ZW (2013) Destructive Tree Diseases Associated with Ambrosia and Bark Beetles: Black Swan Events in Tree Pathology? Plant Disease 97: 856–872. https://doi.org/10.1094/PDIS-01-13-0056-FE

Porras-Alfaro A, Liu K-L, Kuske CR, Xie G (2014) From Genus to Phylum: Large-Subunit and Internal Transcribed Spacer rRNA Operon Regions Show Similar Classification Accuracies Influenced by Database Composition. Applied and Environmental Microbiology 80: 829–840. https://doi.org/10.1128/AEM.02894-13

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41: D590–D596. https://doi.org/10.1093/nar/gks1219

Raman A, Wheatley W, Popay A (2012) Endophytic Fungus-Vascular Plant-Insect Interactions. Environmental Entomology 41: 433–447. https://doi.org/10.1603/EN11317

Richardson RT, Bengtsson-Palme J, Johnson RM (2017) Evaluating and optimizing the performance of software commonly used for the taxonomic classification of DNA metabarcoding sequence data. Molecular Ecology Resources 17: 760–769.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. PeerJ 4: e2584. https://doi.org/10.7717/peerj.2584

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium, Fungal Barcoding Consortium Author List, Bolchacova E, Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An K-D, Bai F-Y, Barreto RW, Begerow D, Bergeron M-J, Blackwell M, Boekhout T, Bogale M, Boonyuen N, Burgaz AR, Buyck B, Cai L, Cai Q, Cardinali G, Chaverri P, Coppins BJ, Crespo A, Cubas P, Cummings C, Damm U, de Beer ZW, de Hoog GS, Del-Prado R, Dentinger B, Dieguez-Uribeondo J, Divakar PK, Douglas B, Duenas M, Duong

TA, Eberhardt U, Edwards JE, Elshahed MS, Fliegerova K, Furtado M, Garcia MA, Ge Z-W, Griffith GW, Griffiths K, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Guo L-D, Hagen F, Hambleton S, Hamelin RC, Hansen K, Harrold P, Heller G, Herrera C, Hirayama K, Hirooka Y, Ho H-M, Hoffmann K, Hofstetter V, Hognabba F, Hollingsworth PM, Hong S-B, Hosaka K, Houbraken J, Hughes K, Huhtinen S, Hyde KD, James T, Johnson EM, Johnson JE, Johnston PR, Jones EBG, Kelly LJ, Kirk PM, Knapp DG, Koljalg U, Kovacs GM, Kurtzman CP, Landvik S, Leavitt SD, Liggenstoffer AS, Liimatainen K, Lombard L, Luangsa-ard JJ, Lumbsch HT, Maganti H, Maharachchikumbura SSN, Martin MP, May TW, McTaggart AR, Methven AS, Meyer W, Moncalvo J-M, Mongkolsamrit S, Nagy LG, Nilsson RH, Niskanen T, Nyilasi I, Okada G, Okane I, Olariaga I, Otte J, Papp T, Park D, Petkovits T, Pino-Bodas R, Quaedvlieg W, Raja HA, Redecker D, Rintoul TL, Ruibal C, Sarmiento-Ramirez JM, Schmitt I, Schussler A, Shearer C, Sotome K, Stefani FOP, Stenroos S, Stielow B, Stockinger H, Suetrong S, Suh S-O, Sung G-H, Suzuki M, Tanaka K, Tedersoo L, Telleria MT, Tretter E, Untereiner WA, Urbina H, Vagvolgyi C, Vialle A, Vu TD, Walther G, Wang Q-M, Wang Y, Weir BS, Weiss M, White MM, Xu J, Yahr R, Yang ZL, Yurkov A, Zamora J-C, Zhang N, Zhuang W-Y, Schindel D (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proceedings of the National Academy of Sciences 109: 6241–6246. https://doi.org/10.1073/pnas.1117018109

Six DL (2020) Niche construction theory can link bark beetle-fungus symbiosis type and colonization behavior to large scale causal chain-effects. Current Opinion in Insect Science 39: 27–34. https://doi.org/10.1016/j.cois.2019.12.005

Skelton J, Loyd A, Smith JA, Blanchette RA, Held BW, Hulcr J (2019) Experimental evidence that fungal symbionts of beetles suppress wood decay by competing with decay fungi. PeerJ Preprints 7:e27676v1. https://doi.org/10.7287/peerj.preprints.27676v1

Stielow JB, Levesque CA, Seifert KA, Meyer W, Iriny L, Smits D, Renfurm R, Verkley G, Groenewald M, Chaduli D (2015) One fungus, which genes? Development and assessment of universal primers for potential secondary fungal DNA barcodes. Persoonia: Molecular Phylogeny and Evolution of Fungi 35: 242. https://doi.org/10.3767/003158515X689135

Tedersoo L, Drenkhan R, Anslan S, Morales-Rodriguez C, Cleary M (2019) High-throughput identification and diagnostics of pathogens and pests: overview and practical recommendations. Molecular ecology resources 19: 47–76.

Tedersoo L, Ramirez KS, Nilsson RH, Kaljuvee A, Kõljalg U, Abarenkov K (2015a) Standardizing metadata and taxonomic identification in metabarcoding studies. GigaScience 4: s13742-015. https://doi.org/10.1186/s13742-015-0074-5

Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson H, Hildebrand F (2015b) Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. MycoKeys 10: 1. https://doi.org/10.3897/mycokeys.10.4852

Tsirogiannis C, Sandel B (2016) PhyloMeasures: a package for computing phylogenetic biodiversity measures and their statistical moments. Ecography 39: 709–714. https://doi.org/10.1111/ecog.01814

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

Vrålstad T (2011) ITS, OTUs and beyond-fungal hyperdiversity calls for supplementary solutions. Molecular Ecology 20: 2873–2875. https://doi.org/10.1111/j.1365-294X.2011.05149.x

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. Applied and Environmental Microbiology 73: 5261–5267. https://doi.org/10.1128/AEM.00062-07

Webb CO, Ackerly DD, Kembel SW (2008) Phylocom: software for the analysis of phylogenetic community structure and trait evolution. Bioinformatics 24. https://doi.org/10.1093/bioinformatics/btn358

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. 2nd ed. 2016. Springer International, Cham, 1 pp. https://doi.org/10.1007/978-3-319-24277-4

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

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring: Biodiversity soup. Methods in Ecology and Evolution 3: 613–623. https://doi.org/10.1111/j.2041-210X.2012.00198.x

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30: 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zhang N, Castlebury LA, Miller AN, Huhndorf SM, Schoch CL, Seifert KA, Rossman AY, Rogers JD, Kohlmeyer J, Volkmann-Kohlmeyer B, Sung G-H (2006) An overview of the systematics of the Sordariomycetes based on a four-gene phylogeny. Mycologia 98: 1076–1087. https://doi.org/10.1080/15572536.2006.11832635