The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species

Eva Lena Estensmo[1], sundy Maurice[1], Luis Morgado[2], Pedro Martin-Sanchez[3], Inger Skrede[1], and Håvard Kauserud[4]

[1]University of Oslo Faculty of Mathematics and Natural Sciences
[2]Naturalis Biodiversity Center
[3]University of Oslo
[4]University in Oslo

September 21, 2020

Abstract

DNA metabarcoding has become a powerful approach for analyzing complex communities from environmental samples, but there are still methodological challenges limiting its full potential. While conserved DNA markers, like 16S and 18S, often are not able to discriminate among closely related species, other more variable markers – like the fungal ITS region, may include considerable intraspecific variation, which can lead to over-splitting of species during DNA metabarcoding analyses. Here we assess the effects of intraspecific sequence variation in DNA metabarcoding, by analyzing local populations of eleven fungal species. We investigated the allelic diversity of ITS2 haplotypes using both Sanger sequencing and high throughput sequencing (HTS), coupled with error correction with the software DADA2. All focal species, except one, included some level of intraspecific variation in the ITS2 region. Overall, we observed a high correspondence between haplotypes generated by Sanger sequencing and HTS, with the exception of a few additional haplotypes detected using either approach. These extra haplotypes, often occurring in low frequencies, were likely due to PCR and sequencing errors or intragenomic variation in the rDNA region. The presence of intraspecific (and possibly intragenomic) variation in ITS2 suggest that haplotypes (or ASVs) should not be used as basic units in ITS-based fungal community analyses, but an extra clustering step is needed to approach species-level resolution.

The influence of intraspecific sequence variation during DNA metabarcoding: A case study of eleven fungal species

Eva Lena F. Estensmo[1], Sundy Maurice[1], Luis Morgado[1,2], Pedro M. Martin-Sanchez[1], Inger Skrede[1], Håvard Kauserud[1]

[1] Section for Genetics and Evolutionary Biology (Evogene), Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316 Oslo, Norway.

[2]Naturalis Biodiversity Center, Vondellaan 55, P.O. Box 9517, 2300, RA Leiden, the Netherlands

*Corresponding authors:

eva.estensmo@outlook.com ; havard.kauserud@ibv.uio.no

Running title: Haplotype diversity impacts DNA metabarcoding

Abstract

DNA metabarcoding has become a powerful approach for analyzing complex communities from environmental samples, but there are still methodological challenges limiting its full potential. While conserved DNA markers, like 16S and 18S, often are not able to discriminate among closely related species, other more variable markers – like the fungal ITS region, may include considerable intraspecific variation, which can lead to over-splitting of species during DNA metabarcoding analyses. Here we assess the effects of intraspecific sequence variation in DNA metabarcoding, by analyzing local populations of eleven fungal species. We investigated the allelic diversity of ITS2 haplotypes using both Sanger sequencing and high throughput sequencing (HTS), coupled with error correction with the software dada2. All focal species, except one, included some level of intraspecific variation in the ITS2 region. Overall, we observed a high correspondence between haplotypes generated by Sanger sequencing and HTS, with the exception of a few additional haplotypes detected using either approach. These extra haplotypes, often occurring in low frequencies, were likely due to PCR and sequencing errors or intragenomic variation in the rDNA region. The presence of intraspecific (and possibly intragenomic) variation in ITS2 suggest that haplotypes (or ASVs) should not be used as basic units in ITS-based fungal community analyses, but an extra clustering step is needed to approach species-level resolution.

Key words: Community ecology; DNA metabarcoding; fungi; ITS, haplotypes

Introduction

High throughput sequencing (HTS) of amplified markers i.e. DNA metabarcoding has become a powerful tool to study microbial communities (Taberlet, et al. 2012; Lindahl, et al. 2013; Goodwin, et al. 2016; Taberlet, et al. 2018). DNA metabarcoding has considerably improved our understanding of the structure and function of microbial communities in different habitats (Tedersoo, et al. 2014; Bahram, et al. 2018), and is also a well-established approach for surveying the biodiversity (Barsoum, et al. 2019) and ecosystem biomonitoring (Douglas, et al. 2012; Stat, et al. 2017).

The commonly used DNA barcoding region for microorganisms lie within the nuclear ribosomal DNA (rDNA). Parts of this region offer conserved primer sites that can be used to amplify broad taxonomic groups, combined with areas of high interspecific and low intraspecific variation in-between, which can provide some degree of taxonomic resolution. The most used rDNA barcoding markers for microorganisms include the Internal Transcribed Spacer (ITS) region for fungi (Nilsson, et al. 2008; Schoch, et al. 2012), the 16S region for bacteria (Stackebrandt and Goebel 1994) and the 18S region for micro-eukaryotes (Hadziavdic, et al. 2014). Due to different evolutionary rates, these markers include contrasting levels of sequence variability and, thus, provide various levels of resolution. In general, the fungal ITS marker includes considerably more sequence variability compared to 18S, and consequently provides higher interspecific resolution, but also some degree of intraspecific variability (Nilsson, et al. 2008; Schoch, et al. 2012).

Although often ignored, the peculiarities of these taxonomic markers imply that the sequences should be processed differently during DNA metabarcoding analyses. For example, in the case of more conserved markers, like 16S and 18S, merging of taxa is a common problem as it underestimates the species diversity, while for the variable ITS marker, splitting of taxa based on intraspecific sequence variation is also a concern in community analyses. In addition, PCR and sequencing errors introduce artificial sequence variation that can be hard to disentangle from naturally occurring intraspecific sequence variability.

A wide array of different bioinformatics approaches has been developed to group and delineate the HTS data into biological entities that are used downstream in community analyses. One first approach was to cluster sequences into operational taxonomic units (OTUs; approximations for biological taxonomic entities), based on a fixed sequence similarity threshold (Schloss, et al. 2009; Caporaso, et al. 2010; Edgar 2013; Westcott and Schloss 2015). Later, more elaborate approaches were developed in order to better distinguish between PCR

and sequencing artefacts and biological sequence variation (Mahe, et al. 2015; Boyer, et al. 2016; Callahan, et al. 2016), and thus, return OTUs better approximating the biological entities.

Although somewhat different solutions have been developed in various software (Pauvert, et al. 2019), a common basic aim in the more recent methods is to identify the underlying haplotypes, present in the template DNA, that gave rise to all the sequence variability generated during PCR and sequencing. In a recent study (Callahan, et al. 2019), it was shown that the software dada2 is able to provide single-nucleotide resolution when analyzing the entire bacterial 16S region. The term *amplicon sequence variants* (ASVs) has been coined for the output of dada2 analyses, which are approximations for the underlying haplotypes. For conserved markers like 16S and 18S, where one single base pair difference can reflect, at least, differences between species and genera, it has been suggested to use ASVs as input for downstream analyses (Callahan, et al. 2017). However, for markers with high level of intraspecific variation, like the ITS marker used for fungi (Nilsson, et al. 2008), this can be highly problematic since the diversity will be tremendously overestimated by treating each ITS haplotype as a biological entity in downstream statistical analyses. Hence, the ASVs will (at best) represent different allelic variants of ITS region, while community ecology is typically based on species-level analyses. To correct for the intraspecific ITS variation, an extra clustering step may be needed to group haplotypes (or ASVs) into species-level OTUs. For fungi and the ITS region, it has been debated at which similarity level sequences should be clustered to approximate the species-level (Caporaso, et al. 2010; Edgar 2013; Westcott and Schloss 2015). Several studies have indicated that 97% represents a reasonable approximation (Nilsson, et al. 2008; Blaalid, et al. 2013). However, such a general threshold might lead to splitting of some taxa and lumping of others (Blaalid, et al. 2013).

Despite the high level of intraspecific ITS sequence variation in fungi, we have little knowledge on how this variation translates into OTU delineation in DNA metabarcoding studies of fungi communities. Here, we assess how DNA metabarcoding, using the fungal ITS2 marker, is able to deal with intraspecific sequence variation, and to what degree this variation leads to over-splitting of taxa. To address this topic, we performed DNA metabarcoding on 176 fungal specimens of 11 basidiomycetes species and compared to the corresponding Sanger sequences. By denoising the sequence data using dada2 (Callahan, et al. 2016), we tested whether the same ITS2 haplotypes were identified by DNA metabarcoding and Sanger sequencing, and to what degree further sequence clustering is needed to approach species-level resolution.

Material and Methods

Eleven wood-decay fungal species (Table 1) were sampled in an old-growth spruce forest in Southeastern Finland (Issakka, Kuhmo). For each species, 16 individual fruit bodies were collected on distinct spruce logs. Given that these fungi typically spread by sexual basidiospores, no clonal dispersal between spruce logs is expected. The fruit body tissue of these fungi is made up of dikaryotic hyphae and heterozygous genotypes are therefore expected if intraspecific ITS2 variation is present (see Fig. S1 for example).

Approximately five square millimeter (mm^2) of tissue were cut out from each fruit body and grinded in 800 µl of 2% CTAB and 1% beta-mercaptoethanol using a Retsch MM200 mixer (4 x 45 s at 25 oscillations). DNA was extracted using a modified CTAB extraction protocol (Murray and Thompson 1980; Gardes and Bruns 1993) and cleaned with the E.Z.N.A Soil DNA kit (Omega Biotek) by adding the HTR reagent and then following the manufacturer´s guidelines. DNA was eluted in 100 µl elution buffer, quantified with Qubit ds DNA BR Assay kit (Life Technologies) and standardized with 10 mM Tris to a concentration range of 5-10 ng/µl.

The 176 fruit body DNA samples were processed into 2 x 96-well PCR plates, together with 16 technical PCR replicates, two identical mock communities (composed of six fungal species with low probability of occurrence in our dataset), and two negative PCR controls. Each library was amplified using a combination of 96 uniquely tagged primers with tags (x) ranging from 7-9 base pairs. The fungal ITS2 region was targeted with the gITS7 (5'-x GTGAR TCATCGAR TCTTTG) (Ihrmark, et al. 2012) and ITS4 (5'-x CTCCGCTTATTGATATG) (White, et al. 1990) primers. The PCR mixture in 25 µl final volume, consisted of 14.6 µl Milli-Q water, 2.5 µl 10x Gold buffer, 0.2 µl dNTP's (25 nM), 1.5 µl reverse and forward

primers (10 µM), 2.5 µl $MgCl_2$ (50 mM), 1.0 µl BSA (20 mg/ml), 0.2 µl AmpliTaq Gold polymerase (5 U/µl) and 5-10 ng/µl of DNA template. The following cycling parameters were used for amplification: enzyme activation at 95 °C for 5 min, followed by 32 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 1 min, and a final extension step at 72 °C for 10 min.

The quality of PCR products was controlled by electrophoresis on a 2% agarose gel prior to normalization using the SequalPrep Normalization Plate Kit (Invitrogen) and eluted in 20 µl elution buffer. The 96 PCR products within each library were pooled, concentrated and purified using Agencourt AMPure XP magnetic beads (Nerliens Meszansky AS) and the DNA concentration was measured with Qubit ds DNA BR Assay kit (Life Technologies). The two libraries were barcoded with Illumina adapters, spiked with 20% PhiX and sequenced in one Illumina MiSeq (Illumina, San Diego, CA, USA) lane with 2×300 base pair paired-end reads at StarSEQ (StarSEQ GmbH, Mainz, Germany).

For comparison, we generated Sanger sequences of ITS2 for the 176 fruiting bodies. Amplification was performed with ITS3 (5'-GCATCGATGAAGAACGCAGC) and ITS4 (5'-TCCTCCGCTTATTGATATGC) primers (White, et al. 1990), with the same PCR mix and program as above. The resulting amplicons were cleaned with ExoProStar (Sigma Aldrich) and sequenced in both directions by Eurofins Genomics (Ebersberg, Germany)

The resulting metabarcoding dataset comprised 25,953,804 reads. The sequences were demultiplexed with cutadapt (Martin 2011) and low quality reads were removed (at least 26 bp overlap between query and target, no indel and minimum length of 100 bp). dada2 (Callahan, et al. 2016) was used for error correction and merging of the reads, without truncating the sequences in order to preserve length variability. Taxonomy was assigned to the raw ASVs by the unite database (Koljalg, et al. 2005) , and the resulting ASV table consisting of 3,647 sequences. For downstream analyses, we retained only 57 sequences assigned to the 11 target species, numerous others occurred in the HTS data because of fungicolous fungi growing inside the fruit bodies.

Both the ASVs and the Sanger sequences were further processed in geneious prime 2020.0.5 (https://www.geneious.com). The Sanger sequences were manually curated and poor-quality sequences were excluded from the dataset. Heterozygous sites were characterized according to the IUPAC nucleotide code, and the forward and reverse reads were merged when possible (depending on quality). Separate sequence alignments were generated from ASVs and Sanger sequences, which were then merged to a single alignment for each species.

The Sanger sequences, many with heterozygous sites due to allelic variability in the dikaryotic tissue, were dephased i.e the consensus sequence of each sample was split into two homozygous sequence strands, and analyzed for DNA polymorphisms in dnasp v.6 (Rozas, et al. 2017). Hence, for each species, we obtained one haplotype dataset from the Sanger sequences and another from the HTS and compared their relative abundance in R (v 3.6.2; R Core Team 2019). Haplotype networks for the 11 species were generated with popart (Leigh and Bryant 2015), displaying the level of intraspecific variation in ITS2. In the calculation of haplotype network, indels were scored as characters, where multi-position gaps were scored as one mutational event. A biplot showing the correspondence in relative abundance of each haplotype across the two datasets was made in R (v 3.6.2; R Core Team 2019). At last, the sequences from the two haplotype datasets were clustered with 97% identity by VSEARCH (Rognes, et al. 2016).

Results

We obtained high quality ITS2 Sanger sequences for 151 out of 176 fruit bodies, ranging from 6 to 16 fruit bodies per species. The remaining fruit bodies either did not amplify or resulted in low-quality sequences, due to fungicolous fungi growing inside the fruit bodies (generating multiple templates) or high level of heterozygosity of indels, leading to chromatograms hard to interpret. The ITS2 sequences were dephased into one to six ITS2 haplotypes per species (Table 1), identifying a total 45 haplotypes from the Sanger dataset. For all species, except *Amylocystis lapponica* represented by haplotype, some level of intraspecific ITS2 sequence variation were present in the local population.

Although we obtained HTS data for 163 out of 176 fruit bodies distributed across the eleven species, for comparative purposes we only focused on the specimens for which Sanger sequences were available. After removing all ITS2 sequences corresponding to fungicolous fungi, a total of 2,316,395 ITS2 sequences were attributed to the 11 target species. After denoising the sequences using dada2 and removing five additional chimeric sequences, we identified between 1 and 8 haplotypes (ASVs) for each species (Table 1), in total 57 haplotypes.

Overall, we detected 65 haplotypes, of which 37 (57%) were shared between the two approaches (Table S1), eight (12.3%) only from Sanger sequencing, while 20 (30.8%) were specific to the HTS data. With some exceptions, a high correspondence was found in the relative abundance of haplotypes across the two datasets (Fig. 1). The haplotype networks (Fig. 2) illustrate the relationship between the haplotypes identified from the two datasets and demonstrate the level of intraspecific variation across species, varying from one haplotype (in*Amylocystis lapponica*) to eleven (in *Phellopilus nigrolimitatus*). The networks also indicate that most haplotypes were closely related, separated by a few mutational steps. Five haplotypes were in very low abundance in the HTS dataset, ten folds lower than what would be expected from a single allele being present in the population (i.e. total read number divided by number of alleles, Table S1). These rare haplotypes likely represent PCR and sequencing errors, or alternatively, intragenomic variation.

After clustering the sequences with 97% identity, we obtained 13 clusters or OTUs for the 11 species. Each species was represented by one OTU, exept for two, *Phellopilus nigrolimitatus* and *Phlebia centrifuga* , which were represented with two OTUs.

Discussion

In general, we observed a good correspondence between the two methods, Sanger sequencing versus DNA metabarcoding, in assessing allelic variation in the ITS2 marker across the eleven fungal species, with 57% of the detected haplotypes shared across the two datasets. We also observed a high correlation in relative abundances of haplotypes across the datasets, where the most striking mismatches were caused by single base pair indels. The additional haplotypes detected by one of the approaches can either be due to methodological errors introduced at various steps, or they may represent *de facto* sequence variation that one of the methods failed to detect.

In some fungal species, intragenomic variation in ITS occurs due to lack of concerted evolution homogenizing the paralogs (Lindner and Banik 2011). Such variation is hard to detect with direct Sanger sequencing, since a consensus sequence is derived from the multiple DNA templates. Although intragenomic ITS paralogs are rare (Lindner, et al. 2013), we cannot rule out the possibility that some of the extra haplotypes detected by HTS represent ITS paralogs.

Alternatively, some of the unique haplotypes appearing in low abundance in the DNA metabarcoding dataset might be due to PCR errors introduced during the initial PCR cycles and that dada2 failed to identify as artifacts. Although dada2 algorithm has a chimeric sequence filter implemented, five chimeric haplotypes occurred in the filtered DNA metabarcoding dataset, with chimeric breakpoints towards either the beginning or the end of the sequences. This exemplifies that a few haplotypes (ASVs), can be erroneous even after dada2 processing. By analyzing full-length 16S rRNA of mock communities of bacteria sequenced with PacBio SCC, a high correspondence was detected between the original templates and the obtained ASVs (Callahan, et al. 2016). However, also in this case, some additional ASVs detected were either due to PCR or sequencing errors, or alternatively, intragenomic 16S variation (Větrovský and Baldrian 2013). It is important to keep in mind that ASVs are probabilistic sequence reconstruction based on error models and thus have an associated uncertainty. When it comes to the additional haplotypes in the Sanger dataset, this could result from erroneous dephasing of the original Sanger sequences.

For all the target species, except one, some level of intraspecific variation in the ITS2 region was detected, even at the fine geographic scale (i.e. a single forest). This corresponds well with the previous literature on intraspecific ITS variability in the fungal kingdom (Smith, et al. 2007; Nilsson, et al. 2008). Nilsson et al. (2008) reported an intraspecific sequence variability of 3.33% (± standard deviation of 5.62) for

Basidiomycota. For some of the target species, sequence variation in the ITS region has been previously reported across broader spatial scales (Kauserud and Schumacher 2002; Kauserud and Schumacher 2003), in line with our results.

It has recently been advocated to use the term ASVs (in our study largely referred to as haplotypes) as the basic units in microbial community analyses (Callahan, et al. 2017). Indeed, this is a reasonable approach for conserved markers, like 16S and 18S, when a single base pair mutation may separate between species or even genera. This is however not the case for variable markers with intraspecific variation. Our results show that in the variable ITS marker, a clustering step is needed after error correction to approach species-level resolution. After clustering our haplotypes, we obtained 13 OTUs representing the 11 species, with two species represented by two haplotypes. The importance of this depends on the study aims. In studies emphasizing beta diversity (community turnover), it has previously been shown that comparable results can be obtained using ASVs or OTUs representing sequence clusters (Glassman and Martiny 2018). In line with this, Botnen et al (2018) demonstrated that beta diversity patterns are highly robust against different clustering levels, ranging from 85% sequence similarity to 99%, both for ITS and 16S data. The most frequent OTUs (or ASVs) drive the community pattern and they largely show the same distributions across different data treatments (Botnen, et al. 2018).

According to our results, we conclude that DNA metabarcoding, based on HTS and error-correction with dada2, to a large extent reflects the allelic variation in natural populations and is a powerful approach to resolve complex communities. Given the limitations of DNA metabarcoding to separate closely related species when targeting single DNA marker, multiple independent DNA markers are often required to improve taxonomic resolution. Yet, we are still not in a position to generate multi-locus datasets from most environmental samples. Alternatively, third generation sequencing technologies (e.g. PacBio, Oxford Nanopore) are promising to generate longer barcodes (e.g. 500-1500 bp for 16S, $>$ 700 bp for ITS and 650 bp for COI) and improve taxonomic resolution (Kennedy, et al. 2018; Tedersoo, et al. 2018). To head towards this direction, we will require databases with higher coverage of different regions to provide more reliable taxonomic classification, in addition to suitable bioinformatics tools.

Acknowledgements

We acknowledge Teppo Helo, Ari Meriruoko from Metsähallitus and Gergely Várkonyi from the Friendship Park Research Center in Kuhmo for help with sampling organization. The research was financially supported by the Research Council of Norway (grant No 254746) and the Department of Biosciences at the University of Oslo.

Competing Interests

The authors declare no competing interest.

Data Accessibility

Sanger sequences are deposited in NCBI. HTS dataset and codes are available at the Dryad Digital Repository (http://dx.doi.org/issakka/its_clustering) and will be made available upon publication.

Author contributions

HK, SM and ELE designed and conceptualized the research. SM and HK performed sampling. SM processed samples and extracted DNA. SM and ELE performed HTS library preparation and Sanger sequencing. ELE analyzed data with contributions from HK, SM, and LM. ELE, SM and HK wrote the manuscript. All authors edited and approved the manuscript.

References

Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H, et al. 2018. Structure and function of the global topsoil microbiome.

Nature 560:233-237.

Barsoum N, Bruce C, Forster J, Ji Y-Q, Yu DW. 2019. The devil is in the detail: metabarcoding of arthropods provides a sensitive measure of biodiversity response to forest stand composition compared with surrogate measures of biodiversity. Ecological Indicators 101:313-323.

Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kauserud H. 2013. ITS1 versus ITS2 as DNA metabarcodes for fungi. Molecular Ecology Resources 13:218-224.

Botnen SS, Davey ML, Halvorsen R, Kauserud H. 2018. Sequence clustering threshold has little effect on the recovery of microbial community structure. Molecular Ecology Resources 18:1064-1076.

Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. 2016. Obitools: a unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources 16:176-182.

Callahan BJ, McMurdie PJ, Holmes SP. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME journal 11:2639-2643.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nature methods 13:581.

Callahan BJ, Wong J, Heiner C, Oh S, Theriot CM, Gulati AS, McGill SK, Dougherty MK. 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Research 47:e103.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature methods 7:335-336.

Douglas WY, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution 3:613-623.

Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature methods 10:996.

Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes–application to the identification of mycorrhizae and rusts. Molecular Ecology 2:113-118.

Glassman SI, Martiny JB. 2018. Broadscale ecological patterns are robust to use of exact sequence variants versus operational taxonomic units. MSphere 3.

Goodwin S, McPherson JD, McCombie WR. 2016. Coming of age: ten years of next-generation sequencing technologies. Nature Reviews Genetics 17:333-351.

Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C. 2014. Characterization of the 18S rRNA gene for designing universal eukaryote specific primers. PLOS ONE 9:e87624.

Ihrmark K, Bödeker I, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandström-Durling M, Clemmensen KE. 2012. New primers to amplify the fungal ITS2 region–evaluation by 454-sequencing of artificial and natural communities. FEMS microbiology ecology 82:666-677.

Kauserud H, Schumacher T. 2002. Population structure of the endangered wood decay fungus Phellinus nigrolimitatus (Basidiomycota). Canadian journal of botany. Journal canadien de botanique 80:597-606.

Kauserud H, Schumacher T. 2003. Regional and local population structure of the pioneer wood-decay fungus Trichaptum abietinum. Mycologia 95:416-425.

Kennedy PG, Cline LC, Song Z. 2018. Probing promise versus performance in longer read fungal metabarcoding. New phytologist 217:973-976.

Koljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Hoiland K, Kjoller R, Larsson E, et al. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New phytologist 166:1063-1068.

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution 6:1110-1116.

Lindahl BD, Nilsson RH, Tedersoo L, Abarenkov K, Carlsen T, Kjoller R, Koljalg U, Pennanen T, Rosendahl S, Stenlid J, 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, Henrik Nilsson R, Davey M, Schumacher T, Kauserud H. 2013. Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. Ecology and Evolution 3:1751-1764.

Mahe F, Rognes T, Quince C, de Vargas C, Dunthorn M. 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ 3:e1420.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17:10-12.

Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8:4321-4325.

Nilsson RH, 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. Evolutionary Bioinformatics 4:EBO.S653.

Pauvert C, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, Lesur I, Vallance J, Vacher C. 2019. Bioinformatics matters: the accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. Fungal ecology 41:23-33.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/index.html.

Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ Preprints 4:e2409v2401.

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology Evolution 34:3299-3302.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75:7537-7541.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. PNAS 109:6241-6246.

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.

Stackebrandt E, Goebel BM. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Evol Microbiol 44:846-849.

Stat M, Huggett MJ, Bernasconi R, DiBattista JD, Berry TE, Newman SJ, Harvey ES, Bunce M. 2017. Ecosystem biomonitoring with eDNA: metabarcoding across the tree of life in a tropical marine environment. Scientific Reports 7:1-11.

Taberlet P, Bonin A, Coissac E, Zinger L. 2018. Environmental DNA: for biodiversity research and monitoring: Oxford University Press.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21:2045-2050.

Tedersoo L, Bahram M, Polme S, Koljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A, et al. 2014. Global diversity and geography of soil fungi. Science 346:1256688.

Tedersoo L, Tooming-Klunderud A, Anslan S. 2018. PacBio metabarcoding of fungi and other eukaryotes: errors, biases and perspectives. New phytologist 217:1370-1385.

Větrovský T, Baldrian P. 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLOS ONE 8:e57923-e57923.

Westcott SL, Schloss PD. 2015. De novo clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. PeerJ 3:e1487.

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 18:315-322.

Tables

Table 1. Comparison of Sanger and HTS sequences. Only specimens for which sequences were obtained from both approaches are shown. Sequence length (base pair) is the overlap between Sanger and HTS sequence alignments, number of dephased sequences correspond to the ITS2 sequence from each dikaryotic (n+n) individual and ASVs stands for amplicon sequence variants. Total haplotypes (Hap.) include common haplotypes from both Sanger and HTS and additional haplotypes identified by either approach.

Species	**Specimen**	**Sequence length (bp)[a]**	**Sanger sequences**	**Sanger sequences**	**Sanger sequences**	**ASVs**	**ASVs**	**ASVs**
			Dephased sequences	Polymorphic sites	Hap.	Reads	Polymorphic sites	Hap.
Amylocystis lapponica	9	308	18	0	1	66,364	0	1
Antrodia serialis	16	188	32	3	4	178,787	3	4
Fomitopsis pinicola	16	229	32	5	5	238,143	7	8
Fomitopsis rosea	15	277	30	5	5	168,991	6	8
Gloeophyllum separium	16	271	32	5	5	429,990	5	5
Phlebia centrifuga	16	276	32	1	2	355,067	10	4

Species	Specimen	Sequence length (bp)[a]	Sanger sequences	Sanger sequences	Sanger sequences	ASVs	ASVs	ASVs
Phellinus ferrugineofuscus	16	282	32	5	4	178,295	5	5
Phellopilus nigrolimitatus	6	296	12	9	6	46,677	8	7
Phellinus viticola	16	281	32	4	5	120,306	4	5
Postia caesia	10	232	20	2	3	191,483	10	4
Trichaptum abietinum	15	268	30	7	5	148,066	7	6

Figures

Figure 1. Biplot showing the correspondence of haplotypes in relative abundance across the two datasets (Sanger sequencing on x-axis and HTS on y-axis).

Figure 2. Haplotype networks displaying the level of intraspecific variation in ITS2 for the 11 fungal species. Each circle represents one haplotype and each dash represent one mutational step. Green color indicates haplotypes detected both by Sanger sequencing and HTS, yellow haplotypes from HTS dataset, and blue haplotypes were only detected by Sanger sequencing. Red arrows indicate haplotypes occurring with very low sequence abundance in the HTS data i.e. 10 times below what would be expected from a single allele in the population. The naming of haplotypes (Hap_1 to Hap_65) follows Table S1.