MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) doi: 10.1111/1755-0998.12065

ITS1 versus ITS2 as DNA metabarcodes for fungi

R. BLAALID,* S. KUMAR,* R. H. NILSSON,† K. ABARENKOV,‡ P. M. KIRK§ and H. KAUSERUD*
*Microbial Evolution Research Group (MERG), Department of Biology, University of Oslo, PO Box 1066 Blindern, Oslo, NO-0316 Norway, †Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden, ‡Natural History Museum, University of Tartu, Tartu, Estonia, §Mycology Section, Royal Botanic Gardens, Kew, UK

Abstract

The nuclear ribosomal Internal Transcribed Spacer ITS region is widely used as a DNA metabarcoding marker to characterize the diversity and composition of fungal communities. In amplicon pyrosequencing studies of fungal diversity, one of the spacers ITS1 or ITS2 of the ITS region is normally used. In this methodological study we evaluate the usability of ITS1 vs. ITS2 as a DNA metabarcoding marker for fungi. We analyse three data sets: two comprising ITS1 and ITS2 sequences of known taxonomic affiliations and a third comprising ITS1 and ITS2 environmental amplicon pyrosequencing data. Clustering analyses of sequences with known taxonomy using the bioinformatics pipeline ClustEx revealed that a 97% similarity cut-off represent a reasonable threshold for estimating the number of known species in the data sets for both ITS1 and ITS2. However, no single threshold value worked well for all fungi at the same time within the curated UNITE database, and we found that the Operational Taxonomic Unit (OTU) concept is not easily translated into the level of species because many species are distributed over several clusters. Clustering analyses of the 134 692 ITS1 and ITS2 pyrosequences using a 97% similarity cut-off revealed a high similarity between the two data sets when it comes to taxonomic coverage. Although some groups are under- or unrepresented in the two data sets due to, e.g. primer mismatches, our results indicate that ITS1 and ITS2 to a large extent yield similar results when used as DNA metabarcodes for fungi.

Keywords: amplicon sequencing, community ecology, DNA barcoding, fungi, internal transcribed spacer (ITS), pyrosequencing

Received 21 June 2012; revision received 15 November 2012; accepted 20 November 2012

Introduction

The internal transcribed spacer region (ITS) of the nuclear ribosomal DNA is the formal DNA barcoding region for molecular identification of fungi (Schoch *et al.* 2012). For nearly two decades the ITS region have been used as an 'unofficial fungal DNA barcode' due to its high degree of interspecific variability, conserved primer sites and multicopy nature in the genome. Since the early 1990s, the ITS region has been heavily used in both molecular systematics and ecological studies of fungi, which has lead to accumulation of more than two hundred thousand Sanger-derived fungal ITS sequences in the international nucleotide sequence databases (INSD: GenBank, ENA and DDBJ; Karsch-Mizrachi *et al.* 2012). The fungal ITS region varies roughly, with some exceptions, between approximately 450 and 750 base pairs (bp) in length and consists of three subregions: the variable spacers ITS1 and ITS2 and the intercalary 5.8S gene. While the 5.8S gene is highly conserved, the ITS1 and ITS2 spacers normally provide resolution at a within-genus and often within-species level (Nilsson *et al.* 2008). Although the ITS region has been selected as the fungal DNA barcode, it lacks the necessary resolution in some groups of fungi (e.g. Gazis *et al.* 2011). In groups such as *Fusarium* and *Aspergillus*, other or additional markers are necessary (Balajee *et al.* 2009).

During the last few years, fungal ecology has started to draw advantages from technological achievements in high-throughput DNA sequencing where approaches such as 454 pyrosequencing (Margulies *et al.* 2005) can generate hundreds of thousands of DNA sequences from any given sample in the course of a day (Buèe *et al.* 2009; Jumpponen & Jones 2009; Tedersoo *et al.* 2010; Wallander *et al.* 2010; Blaalid *et al.* 2012; Davey *et al.* 2012). High-throughput DNA-based identification of multiple species from a single complex environmental sample was recently defined as DNA metabarcoding (Taberlet *et al.* 2012), a term adopted here. Due to restrictions in sequence read lengths, only a part of the ITS region is normally analysed in amplicon pyrosequencing

Correspondence: Håvard Kauserud, Fax: 47 22854726;
E-mail: haavarka@bio.uio.no

studies—either ITS1 (Buèe *et al.* 2009; Jumpponen & Jones 2009; Tedersoo *et al.* 2010; Blaalid *et al.* 2012) or ITS2 (Wallander *et al.* 2010; Davey *et al.* 2012). With improvements of the sequencing technologies, it will be possible to sequence the entire ITS region. However, as chimeric sequences typically are formed across the conserved 5.8S regions during PCR amplification of mixed substrates, it may still be favourable to analyse ITS1 and ITS2 separately.

Whether ITS1 or ITS2 provides the best taxonomic resolution at the species level has been debated (Nilsson *et al.* 2008; Bellemain *et al.* 2010; Mello *et al.* 2011). Furthermore, different types of primer biases may lead to biased amplification of different taxonomic groups in ITS1 and ITS2 (Tedersoo *et al.* 2010). Systematic length differences between ITS1 and ITS2 may have a similar effect (Bellemain *et al.* 2010). It has been shown that ITS1 is on average somewhat more variable than ITS2 in most fungal lineages (Nilsson *et al.* 2008; Ryberg *et al.* 2008; Mullineux & Hausner 2009). An evaluation of 4185 complete ITS reads revealed that 66% had a higher variability in ITS1, whereas 34% were more variable in ITS2 (Nilsson *et al.* 2008). Separate studies of members within both Basidiomycota, such as the genus *Inocybe*, and Ascomycota, such as the order Ophiostomatales, indicate that ITS1 is more variable compared with ITS2 (Ryberg *et al.* 2008; Mullineux & Hausner 2009). Although there are species differences between the variability of ITS1 vs. ITS2, a correlation of this variability has been observed, suggesting that the two regions do not evolve separately (Nilsson *et al.* 2008). In a recent pyrosequencing study, the ability of ITS1 and ITS2 to uncover soil fungal communities in truffle grounds was compared (Mello *et al.* 2011). The authors concluded that both ITS1 and ITS2 were suitable as DNA metabarcoding markers (Mello *et al.* 2011).

In this study we use both sequence data of known taxonomic affiliation as well as pyrosequencing data obtained from 42 environmental samples to evaluate the usability of ITS1 vs. ITS2 as DNA metabarcodes for fungi. Although the usability of ITS1 vs. ITS2 has been addressed in other studies (Mello *et al.* 2011), to our knowledge this is the first attempt to examine how well clustering techniques group ITS sequences of known taxonomic affiliation into species, as judged by Latin binomials.

Materials and methods

Sequence data of known taxonomic affiliation from UNITE and INSD

The first data set consisted of 1743 full-length ITS sequences, comprising 750 unique species accessioned in the fungal reference database UNITE. This is a curated database whose core data are made up by high-quality sequences identified by taxonomic experts (Abarenkov *et al.* 2010a). Most (>95%) of the analysed accessions in UNITE were basidiomycetes. The other data set comprised ITS data with known taxonomy obtained from INSD. First, all available fungal ITS sequences with a full Latin binomial were downloaded from INSD. Reverse complementary entries were corrected for; sequences of poor read quality, as well as chimeric sequences, were pursued for exclusion through PlutoF (Abarenkov *et al.* 2010b). Only the 26 665 full-length sequences belonging to Dikarya (Ascomycota + Basidiomycota) were used in the further analyses. The sequences were split into one ascomycete data set (18 215 sequences) and another basidiomycete data set (8440 sequences). The ITS data sets were split into ITS1 and ITS2 parts through the ITS Extractor (Nilsson *et al.* 2010). We tested the effect of different sequence similarity levels during clustering with BLASTCLUST using a newly designed bioinformatics pipeline ClustEx (see below). All data sets were clustered using 90–100% sequence similarity (e.g. 90%, 91%, … 100%) requiring 75% sequence coverage. The analysis of the distribution of species across clusters was based on the sequences downloaded from UNITE, including the ratio between known species vs. number of clusters, using Perl scripts.

Software operation

The bioinformatics pipeline ClustEx, used to analyse the UNITE and INSD data, was written in Perl v5.8 and PHP 4.3 as a graphical-interface wrapper for the three clustering programs BLASTCLUST, CD-HIT and UCLUST (Altschul *et al.* 1997; Li & Godzik 2006; Edgar 2010). In addition to user specification of similarity—and coverage range to compare in these programs—ClustEx also provides a few pre- and postclustering functionalities (Fig. S1). Sequences containing (i) DNA ambiguity symbols; (ii) user-defined short sequences (e.g. < 50 bp) and (iii) identical sequences can be removed in any combination of the three prior to clustering. Postclustering functionalities include the useful option to exclude singletons, hence reducing the impact of PCR and sequencing errors (Tedersoo *et al.* 2010). To summarize the effect of different clustering settings, column and line graphs are generated through Perl and R scripts that are integrated in the program and presented to the user. ClustEx produces several output files based on the parameters and clustering method selected. It is implemented at the web-based bioinformatics service Bioportal at the University of Oslo (Kumar *et al.* 2009). Bioportal is freely available to academic users at http://www.bioportal.uio.no/.

ITS1 and ITS2 pyrosequences from environmental samples

Hundred and twenty specimens, including the roots, of the ectomycorrhizal plant *Bistorta vivipara* were collected in a 2 × 2 m plot at Finse, an alpine region in south Norway (60°58′N, 7°31′E). The plant roots were washed and rinsed prior to DNA extraction. The plot was split into 64 subplots and one soil sample (ca. 28 cm^3) was obtained from the centre of each subplot. DNA was extracted from the plant roots using a CTAB extraction protocol as described in previous studies (Blaalid *et al.* 2012; Kauserud *et al.* 2012). The DNA extracts were further cleaned using E.Z.N.A soil clean-up kit (Omega Biotek) following the manufacturer's protocol. Total genomic DNA in soil samples was isolated from 2 ml of homogenized soil slurry using the MO BIO soil extraction kit (MO BIO Laboratories Inc.). Both plant roots and soil samples were prepared for 454 pyrosequencing by performing nested PCR, as recommended by Berry *et al.* (2011), to avoid the ligation bias known to occur in barcode primer-free approaches (Gillevet *et al.* 2010). Highly replicable results have earlier been obtained using this set-up (Kauserud *et al.* 2012). The fungal-specific primers ITS1-F and ITS4 (White *et al.* 1990; Gardes & Bruns 1993) were used in the first PCR whereas the primer pairs ITS2/ITS5 and ITS3/ITS4 were used for amplifying ITS1 and ITS2 respectively. To ensure sample recognition in downstream analyses, different 10 bp tags were added to the primer pairs ITS2/ITS5 and ITS3/ITS4 together with the 454 sequencing adaptors A and B. PCR was performed in 20 μL reaction volumes containing 2 μL template DNA and 18 μL reaction mix. Final concentrations in the PCR were 0.16 mM dNTP mix, 0.2 μM of each primer and 0.4 units Finnzymes Phusion polymerase. The amplification program for both steps of the nested PCR was as follows: 30 s at 98 °C, followed by 19 cycles of 10 s at 98 °C, 20 s at 55 °C, 20 s at 72 °C and a final extension step at 72 °C for 7 min before storage at 4 °C. A high number of PCR cycles were used to reliably obtain successful amplification, as evaluated by gel electrophoresis. A high number of cycles may, however, distort the proportion of various taxa during PCR, and, hence, the sequence data were only analysed qualitatively as presence/absence of OTUs (i.e. not quantitatively). PCR products were cleaned with Wizard© SV Gel and PCR Clean-Up System (Promega), quantified using a Sequalprep Normalization kit (Invitrogen) and pooled into 16 equimolar amplicon libraries. The amplicon libraries were sequenced on a full plate split into 16 lanes using the Roche GS FLX Titanium Series 454 sequencing platform at the Norwegian High-Throughput Sequencing Centre. Both the first and the nested PCR reaction were conducted independently for amplifying ITS1 and ITS2. The raw data were accessioned in the NCBI Sequence Read Archive (Kodama *et al.* 2012), accession no. SRA050547.

We analysed the raw data by filtering and clustering of the ITS1 and the ITS2 regions separately using the bioinformatics pipeline CLOTU (Kumar *et al.* 2011). Reads with sequencing errors in the forward primer and tag, as well as noncongruent tag combinations, were removed. Short reads (<150 bp) and low-quality reads (all reads including one or more N) were filtered out. The sequences were then clustered into operational taxonomic units (OTUs) as a crude approximation of species, using the BLASTCLUST single linkage clustering approach with the requirement of 75% overlap in sequence length between reads in the pairwise alignments and 97% sequence similarity. To reduce putative effects of low sequencing depth, only samples with >1000 reads for both of ITS1 and ITS2 were retained for further analyses, which included 29 plant root samples and 13 soil samples. For both the ITS1 and ITS2 data sets, all clusters containing five sequences or less within individual samples were removed from downstream analyses to reduce potential effects of sequencing errors and tag-switching artefacts (Quince *et al.* 2009; Tedersoo *et al.* 2010; Carlsen *et al.* 2012).

For taxonomic annotation, one representative sequence from each OTU was used in BLASTn searches (Altschul *et al.* 1997) against the NCBI-nr database. Congruency among the top hits was assessed and a higher taxonomic level reported in doubtful cases.

Results and discussion

Analyses of sequences data of known taxonomic affiliation

Clustering analyses at various similarity thresholds yielded largely similar results between ITS1 and ITS2 when it comes to the UNITE data set (Fig. 1a), that largely includes basidiomycetes, and the INSD basidiomycete data set (Fig. 1b). A higher degree of discrepancy was observed between ITS1 and ITS2 in the INSD ascomycete data set (Fig. 1c), where ITS1 generally yielded more clusters than ITS2. This effect was especially pronounced at lower sequence similarity levels (<97%). The results were highly similar for the three clustering approaches BLASTCLUST, CD-HIT and UCLUST (results not shown). Furthermore, the degree of sequence coverage during clustering had little impact on the number of clusters obtained in the range 60–90% coverage. Our results (Fig. 1) indicate that clustering at a 97% sequence similarity level gives a reasonable approximation of the number of known species in the basidiomycete data sets (judged as Latin binomials). In the INSD ascomycete

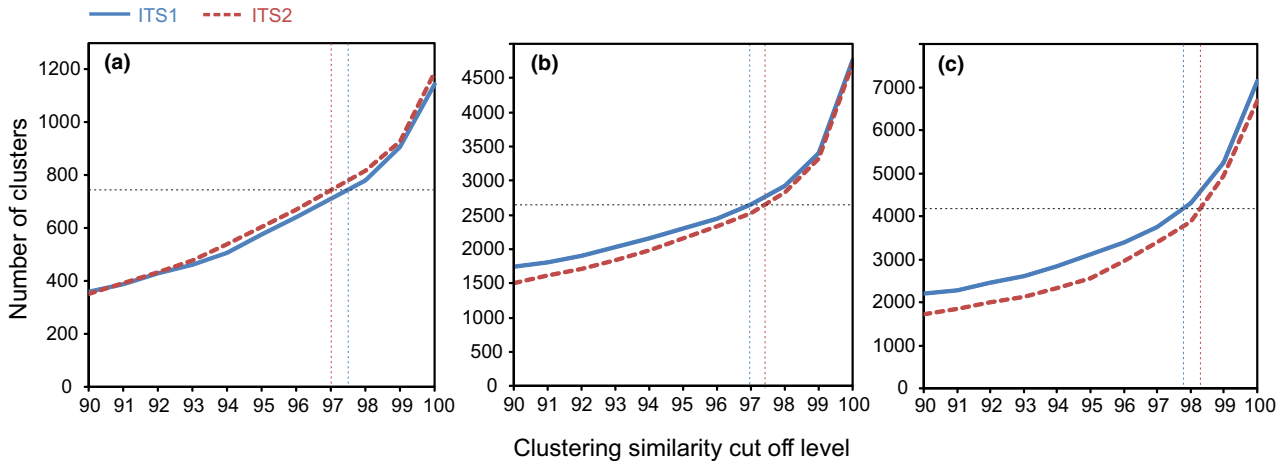

Fig. 1 Number of clusters plotted against each percentage identity cut-off (90–100% similarity) using the ITS1 and ITS2 sequences from (a) UNITE, (b) INSD basidiomycete data, (c) INSD ascomycete data. Stippled horizontal lines indicate the number of known species in the data sets whereas the vertical stippled lines indicate the corresponding level of sequence similarity in the ITS1 and ITS2 data sets.

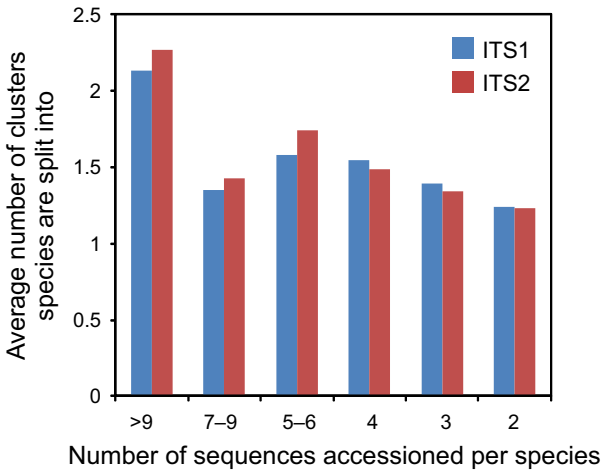

Fig. 2 Histogram showing the average number of clusters species (judged as Latin binomials) are split into (*Y*-axis) according to number of sequences accessioned for each species in the UNITE database (*X*-axis). Hence, species represented by more than nine sequences in UNITE are on average split into approximately two clusters. 97% sequence similarity was used as cut-off during clustering with BLASTCLUST.

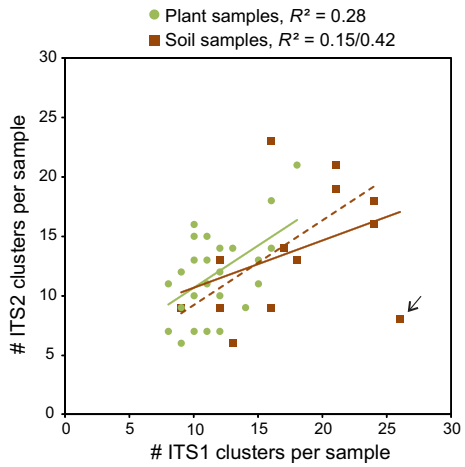

Fig. 3 Linear regression of the relationship between the number of clusters recovered from clustering the ITS1 and ITS2 sequences obtained from the analyses of plant root (circle) and soil (square) samples. The stippled regression line excluded the outlier soil sample indicated by the arrow.

data set, 98% sequence similarity gave a better approximation. However, this could be due to the inclusion of synonyms and telemorphs/anamorphs in this data set, which were difficult to control for. The '3% golden rule' of sequence dissimilarity to demarcate distinct species was devised for full-length ribosomal small-subunit (SSU) sequences of prokaryotes (cf. Stackebrandt & Goebel 1994), and it has been adopted somewhat blindly also in situations where the sequences at hand are not full length, where they are not of prokaryote origin and where they do not represent the SSU in the first place. The fungal ITS region is such an example. The present results nevertheless lend support to the numerous analyses where a 97% sequence similarity level has been used (O'Brien *et al*. 2005; Morris *et al*. 2008; Ryberg *et al*. 2008; Walker *et al*. 2008; Bjorbækmo *et al*. 2010; Tedersoo *et al*. 2010), supporting the similar observation, although with specific focus on basidiomycetes, by Tedersoo *et al*. (2010).

When we inspected the obtained clusters of the UNITE data set more in depth, we observed that 30.6% (ITS1) and 29.2% (ITS2) of the species (Latin binomials) represented by more than two sequences were split into two or more clusters (Fig. 2) and that many clusters

(29.5% for ITS1 and 29.7% for ITS2) featured multiple species (Latin binomials). Hence, there were no major differences in how ITS1 and ITS2 performed when it comes to splitting of sequences of the same species across different clusters. It is not surprising that species are split or lumped into different clusters as there is no biological reason why there would be a single threshold value to demarcate intra- and interspecific variability across the fungal tree of life. On the contrary, the last few years have brought plenty of data to show that many lineages of the fungal kingdom differ in terms of distance among species (Nilsson *et al.* 2008; Gazis *et al.* 2011). New clustering algorithms, not primarily based on direct sequence similarity, offer promise in this regard (Pommier *et al.* 2009; Zinger *et al.* 2009).

Environmental ITS1 and ITS2 pyrosequence data

A total of 134 872 pyrosequences were retained after filtering: 66 317 ITS1 and 68 555 ITS2 sequences. Clustering at 97% sequence similarity followed by filtering of low-frequency clusters (see above) resulted in 116 (ITS1) and 133 (ITS2) clusters respectively. Notably, the average BLAST similarity matches were significantly higher for ITS2 compared with ITS1 (T-test, $P = 0.043$), probably indicating that more ITS2 data are accessioned in the INSD databases compared with ITS1 (Nilsson *et al.* 2009). In some phylogenetic studies (Larsson *et al.* 2004; Miller *et al.* 2006), the ITS2 + a part of the large subunit region has been sequenced without including ITS1. Reassuringly, a significant relationship was observed between the number of clusters obtained with ITS1 and ITS2 (Fig. 3).

At the phylum level, the taxonomic composition in the ITS1 and ITS2 data sets was roughly similar, with 44% and 34% of the OTUs ascribed to Ascomycota and Basidiomycota for ITS1 and 47% and 31% as corresponding numbers for ITS2. The small difference at the phylum level (3%) may be a result of primer bias (Bellemain *et al.* 2010). Both markers identified 10–11% of the OTUs as Zygomycota s.l. (Fig. 4). At the order level, the most dominant basidiomycete groups in both the ITS1 and ITS2 data sets were Agaricales, Thelephorales, Sebacinales and Tremellales (Fig. 4). These were reported in roughly equal abundances with ITS1 and ITS2. Helotiales was the most abundant ascomycete order, comprising 19% of the OTUs for both ITS1 and ITS2. When it comes to number of OTUs, Pezizales were slightly more abundant in the ITS1 data set.

Although ITS1 and ITS2 in general yielded a similar taxonomic composition at the phylum and order level, some differences were noted. For example, Geoglossales was absent in the ITS1 data set. However, only two OTUs were included in this group. Hence, PCR stochasticity cannot be excluded as a cause for this observation. Furthermore, Lecanorales were more abundant in the ITS2 data set. Members of this order are known to have introns at the end of the rDNA small subunit (SSU) and may therefore be more difficult to amplify using, e.g. ITS5. Within Agaricales a higher number of genera appeared in the ITS1 data set whereas higher species diversity within *Cortinarius* was detected in the ITS2 data set. As above, these discrepancies might be due to biased amplification. However, differences in taxonomic resolution at a genus level between ITS1 and ITS2, as well as different amounts of reference sequences might also account for the observed differences.

Our results largely corroborate earlier observations that structural similarities of the fungal communities are observed across ITS1 and ITS2 (Jumpponen 2004; Arfi *et al.* 2011; Mello *et al.* 2011). However, direct comparisons across studies are difficult due to differences in data treatment. While we filtered out low-frequency clusters to remove PCR and sequencing artefacts (Tedersoo *et al.*

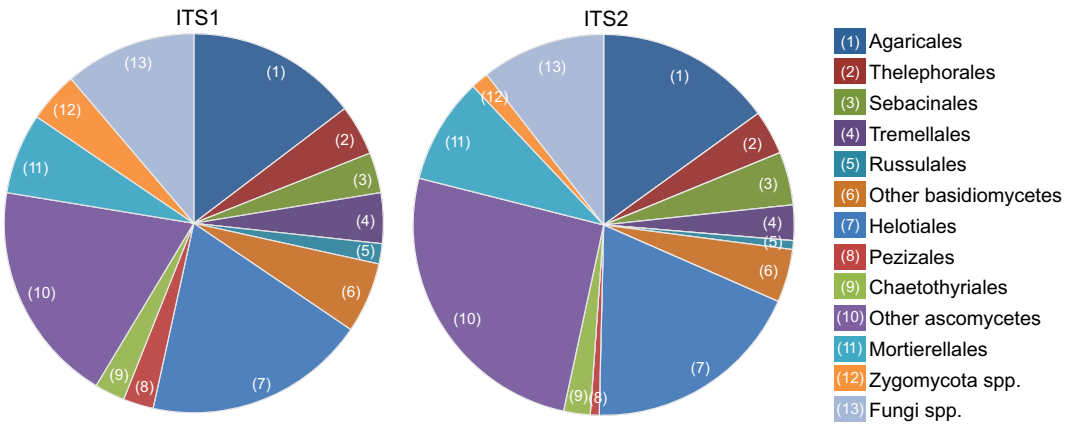

Fig. 4 Taxonomic composition recovered in the environmental samples using ITS1 (116 clusters) and ITS2 (133 clusters) as DNA metabarcodes respectively.

2010), singletons were retained in Mello *et al.* (2011). This study highlights that ITS1 and ITS2 to a large extent yield similar results when used as DNA metabarcodes for fungi. Furthermore, our analyses demonstrate the limitations of ITS as a marker, as ITS sequences from different species often cluster together and many species split into several clusters.

Acknowledgements

We acknowledge Balder Holm for assistance during fieldwork and the staff at Finse Alpine Research station for providing lab facilities. Xiaoli Zhang is acknowledged for help with programming. Tor Carlsen is acknowledged for supervision and valuable discussions on 454 sequencing, The Norwegian Sequencing Centre for 454 pyrosequencing, and the Bioportal at the University of Oslo for providing bioinformatics facilities. RHN acknowledges financial support from FORMAS (215-2011-498) and from the Carl Stenholm Foundation.

References

Abarenkov K, Nilsson RH, Larsson KH *et al.* (2010a) The UNITE database for molecular identification of fungi–recent updates and future perspectives. *New Phytologist*, **186**, 281–285.

Abarenkov K, Tedersoo L, Nilsson RH *et al.* (2010b) PlutoF–a Web based workbench for ecological and taxonomical research, with an online implementation for fungal ITS sequences. *Evolutionary Bioinformatics*, **6**, 189–196.

Altschul SF, Madden TL, Schäffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Arfi Y, Buée M, Marchand C, Record E (2011) Multiple markers pyrosequencing reveals highly diverse and host-specific fungal communities on the mangrove trees *Avicennia marina* and *Rhizophora stylosa*. *FEMS Microbiology Ecology*, **79**, 433–444.

Balajee SA, Borman AM, Brandt ME *et al.* (2009) Sequence-based identification of *Aspergillus*, *Fusarium*, and *Mucorales* species in the clinical mycology laboratory: where are we and where should we go from here? *Journal of Clinical Microbiology*, **47**, 877–884.

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

Berry D, Ben K, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, **77**, 7846–7849.

Bjorbækmo MFM, Carlsen T, Brysting AK *et al.* (2010) High diversity of root associated fungi in both arctic and alpine *Dryas octopetala*. *BMC Plant Biology*, **10**, 244.

Blaalid R, Carlsen T, Kumar S *et al.* (2012) Changes in the root-associated fungal communities along a primary succession gradient analysed by 454 pyrosequencing. *Molecular Ecology*, **21**, 1897–1908.

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

Carlsen T, Aas AB, Lindner D *et al.* (2012) Don't make a mista(g)ke: Is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecology*, **5**, 747–749.

Davey ML, Heegaard E, Halvorsen R *et al.* (2012) Seasonal trends in the biomass and structure of bryophyte-associated fungal communities explored by 454-pyrosequencing. *New Phytologist*, **195**, 844–856.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.

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.

Gazis R, Rhener S, Chaverri P (2011) Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Molecular Ecology*, **30**, 3001–3013.

Gillevet P, Sikaroodia M, Keshavarzianc A, Mutluc EA (2010) Quantitative Assessment of the Human Gut Microbiome Using Multitag Pyrosequencing. *Chemistry & Biodiversity*, **7**, 1065–1075.

Jumpponen A (2004) Soil fungal communities underneath willow canopies on a primary successional glacier forefront: rDNA sequence results can be affected by primer selection and chimeric data. *Microbial Ecology*, **53**, 233–246.

Jumpponen A, Jones KL (2009) Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phylloshere. *New Phytologist*, **184**, 438–448.

Karsch-Mizrachi I, Nakamura Y, Cochrane G *et al.* (2012) The International Nucleotide Database Collaboration. *Nucleic Acids Research*, **40**, 33–37.

Kauserud H, Kumar S, Brysting AK *et al.* (2012) High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples. *Mycorrhiza*, **22**, 309–315.

Kodama Y, Shumway M *et al.* (2012) The sequence read archive: explosive growth of sequence data. *Nucleic Acids Research*, **40**, 54–56.

Kumar S, Skjærveland Å, Orr R *et al.* (2009) AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. *BMC Bioinformatics*, **10**, 357.

Kumar S, Carlsen T, Mevik B *et al.* (2011) CLOTU: an online pipeline for processing and clustering of 454 amplicon reads into OTUs followed by taxonomic annotation. *BMC Bioinformatics*, **12**, 182.

Larsson KH, Larsson E, Kõljalg U (2004) High phylogenetic diversity among corticioid homobasidiomycetes. *Mycological Research*, **108**, 983–1002.

Li W, Godzik A (2006) CD-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Mello A, Napoli C, Morin CME *et al.* (2011) ITS-1 versus ITS-2 pyrosequencing: a comparison of fungal populations in truffle grounds. *Mycologia*, **103**, 1184–1193.

Miller SL, Larsson E, Larsson KH *et al.* (2006) Perspectives in the new Russulales. *Mycologia*, **98**, 960–970.

Morris HM, Smith ME, Rizzo DM *et al.* (2008) Contrasting ectomycorrhizal fungal communities on the roots of co-occurring oaks (*Quercus spp.*) in a California woodland. *New Phytologist*, **178**, 167–176.

Mullineux T, Hausner G (2009) Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera *Grosmannia* and *Leptographium*. *Fungal Genetics and Biology*, **46**, 855–867.

Nilsson RH, Kristiansson E, Ryberg M *et al.* (2008) Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. *Evolutionary Bioinformatics*, **4**, 193–201.

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

Nilsson RH, Veldre V, Hartmann M *et al.* (2010) An open source package for automated extraction of ITS1 and ITS2 from fungal ITS sequences for use in high-troughput community assays and molecular ecology. *Fungal Ecology*, **3**, 284–287.

O'Brien HE, Parrent JL, Jackson JA *et al.* (2005) Fungal Community analysis by large-scale sequencing of environmental samples. *Applied and Environmental Microbiology*, **71**, 5544–5550.

Pommier T, Cranbäck B, Lundberg P *et al.* (2009) RAMI: a tool for identification and characterization of phylogenetic clusters in microbial communities. *Bioinformatics*, **25**, 736–742.

Quince C, Lanzén A, Curtis TP *et al*. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods*, **6**, 639.

Ryberg M, Kristiansson E, Sjökvist E, Nilsson RH (2008) An outlook on the fungal internal transcribed spacer sequences in GenBank and the introduction of a web-based tool for the exploration of fungal diversity. *New Phytologist*, **181**, 471–477.

Schoch CL, Seifert K, Huhndorf S *et al.* (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 6241–6246.

Stackebrandt E, Goebel BM (1994) Taxonomic Note: A Place for DNA-DNA Re-association and 16s rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic Bacteriology*, **44**, 846–849.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. *Molecular Ecology*, **21**, 1789–1793.

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

Walker JF, Miller OK Jr, Horton JL (2008) Seasonal dynamics of ectomycorrhizal fungus assemblages on oak seedlings in the southeastern Appalachian Mountains. *Mycorrhiza*, **18**, 123–132.

Wallander H, Johansson E, Brandström M, Lindahl B (2010) Production of ectomycorrhizal mycelium peaks during canopy closure in Norway spruce forests. *New Phytologist*, **187**, 1124–1134.

White TJ, Bruns T, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: A Guide to Methods and Applications*, pp. 315–322. Academic Press, Orlando, Florida.

Zinger L, Gobet A, Pommier T (2009) Two decades of describing the unseen majority of aquatic microbial diversity. *Molecular Ecology*, **21**, 1878–1896.

H.K. and R.B. designed the study, R.B. did field work and molecular analyses, R.H.N., K.A. and P.K. provided data, S.K. did programming and bioinformatics, R.H.N., R.B. and H.K. wrote the manuscript and S.K., K.A. and P.K. provided comments.

Data Accessibility

DNA sequences: NCBI SRA: SRA050547, DRYAD entry doi: 10.5061/dryad.k37m7

Sampling location: Finse, Central Norway, (60°58′N, 7°31′E).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Screenshot of ClustEx main window. The parameters shown include the similarity threshold and coverage ranges as well as several options for sequence filtering.