

Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi

Leho Tedersoo¹, Sten Anslan², Mohammad Bahram^{2,3}, Sergei Põlme^{1,2}, Taavi Riit², Ingrid Liiv², Urmas Kõljalg³, Veljo Kisand⁴, R. Henrik Nilsson⁵, Falk Hildebrand⁶, Peer Bork⁶, Kessy Abarenkov¹

1 Natural History Museum, University of Tartu, Tartu, Estonia **2** Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia **3** Department of Organismal Biology, Uppsala University, Uppsala, Sweden **4** Institute of Technology, Tartu University, Nooruse 1, Tartu 50411, Estonia **5** Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden **6** European Molecular Biology Laboratory, Heidelberg, Germany

Corresponding author: *Leho Tedersoo* (leho.tedersoo@ut.ee)

Academic editor: *T. Lumbsch* | Received 9 March 2015 | Accepted 29 April 2015 | Published 13 May 2015

Citation: Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson RH, Hildebrand F, Bork P, Abarenkov K (2015) Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycKeys* 10: 1–43. doi: 10.3897/mycokeys.10.4852

Abstract

Rapid development of high-throughput (HTS) molecular identification methods has revolutionized our knowledge about taxonomic diversity and ecology of fungi. However, PCR-based methods exhibit multiple technical shortcomings that may bias our understanding of the fungal kingdom. This study was initiated to quantify potential biases in fungal community ecology by comparing the relative performance of amplicon-free shotgun metagenomics and amplicons of nine primer pairs over seven nuclear ribosomal DNA (rDNA) regions often used in metabarcoding analyses. The internal transcribed spacer (ITS) barcodes ITS1 and ITS2 provided greater taxonomic and functional resolution and richness of operational taxonomic units (OTUs) at the 97% similarity threshold compared to barcodes located within the ribosomal small subunit (SSU) and large subunit (LSU) genes. All barcode-primer pair combinations provided consistent results in ranking taxonomic richness and recovering the importance of floristic variables in driving fungal community composition in soils of Papua New Guinea. The choice of forward primer explained up to 2.0% of the variation in OTU-level analysis of the ITS1 and ITS2 barcode data sets. Across the whole data set, barcode-primer pair combination explained 37.6–38.1% of the variation, which surpassed any environmental signal. Overall, the metagenomics data set recovered a similar taxonomic overview, but resulted in much lower fungal rDNA sequencing depth, inability to infer OTUs, and high

uncertainty in identification. We recommend the use of ITS2 or the whole ITS region for metabarcoding and we advocate careful choice of primer pairs in consideration of the relative proportion of fungal DNA and expected dominant groups.

Key words

High-throughput sequencing, internal transcribed spacer (ITS), nuclear large subunit (LSU), nuclear small subunit (SSU), Illumina MiSeq sequencing, shotgun metagenomics, primer bias, taxonomic coverage, identification bias

Introduction

Fungi are one of the most diverse kingdoms of life on Earth (Blackwell 2011). Traditionally, fungi have been identified based on morphological characters of fruit-bodies or pure cultures in agar medium. Because most fungal species do not seem to form conspicuous fruit-bodies or are difficult to isolate into culture, molecular methods have been widely implemented to shed light onto the occurrence and ecological patterns of these fascinating organisms. In particular, high-throughput sequencing (HTS) technologies enable documentation and characterization of hundreds to thousands of species simultaneously from biological samples. These methods have been used successfully to uncover diversity and community composition of fungi in various substrates and localities, including little-studied habitats such as mangroves, aquatic ecosystems, and arctic soils (Arfi et al. 2012; Blaaliid et al. 2013; Livermore and Mattes 2013). Similarly to other amplicon-based methods, HTS techniques are sensitive to PCR and primer biases (Lindahl et al. 2013). At least in theory, ultra-HTS technologies such as Illumina HiSeq may provide sufficient sequencing depth to detect barcode sequences from the entire metagenome as demonstrated for prokaryotes and soil animals (Zhou et al. 2013; Logares et al. 2014). The genomes of eukaryotes are typically several orders of magnitude larger compared to bacterial genomes such that the relative proportion of ribosomal DNA in the metagenome will be lower. This may complicate their detection without enrichment.

The internal transcribed spacer (ITS) region of the nuclear ribosomal DNA (rDNA) is the formal barcode for molecular identification of fungi (Schoch et al. 2012). Alternative ribosomal DNA markers located in the nuclear rDNA large subunit (LSU/28S) and small subunit (SSU/18S) genes are often used to address phylogenetic diversity of fungi, especially in groups such as the former Zygomycota and Chytridiomycota (the so-called early diverging lineages) and Glomeromycota. Taxa belonging to the little-studied phyla Cryptomycota and Chytridiomycota as well as groups comprising mostly unicellular species are difficult to place phylogenetically based on variable markers such as the ITS region due to high genetic divergence and paucity of reference sequences (Ishii et al. 2015). Because taxonomic resolution strongly depends on the choice of barcodes (Arfi et al. 2012; Schoch et al. 2012; Kohout et al. 2014), the use of different markers and analysis methods renders studies on fungal biodiversity

largely incomparable. In addition, alternative primers targeting the same barcode are known to differ in the recovery of taxa both *in silico* (Bellemain et al. 2010) and *in vivo* (Tedersoo et al. 2010; Op de Beeck et al. 2014).

Discussion related to potential taxonomic biases in relation to the class Archaeorhizomycetes due to the choice of primers in our global soil analysis (Tedersoo et al. 2014; Schadt and Rosling *in press*; C.W. Schadt and A. Rosling, pers. comm.) motivated us to undertake this study. Here we address potential biases in diversity and taxonomic composition of fungi in relation to PCR-free approaches and barcode and primer pair combinations of amplicons. We specifically aim to answer the following questions: i) do amplicon-free and PCR-based techniques reveal comparable results?; ii) are there substantial primer biases?; iii) to what extent does the recovered taxonomic richness differ among selected barcode-primer pair combinations?; and iv) do markers differ in their ability to recover ecological patterns in community-level analyses? This study seeks to evaluate the suitability of amplicon-free HTS technologies in fungal metabarcoding and to elaborate on the most suitable barcodes and primers for amplicon-based HTS studies by individual research groups and global metabarcoding consortia based on *in vivo* and *in silico* analyses.

Methods

Sampling and molecular analysis

Between 5 and 30 November 2011, 34 composite soil samples were collected from woody plant-dominated ecosystems in Papua New Guinea (PNG) following a standard protocol (Tedersoo et al. 2014). Briefly, two soil cores (5 cm diam. to 5 cm depth) were collected at a distance of 1-1.5 m from 20 randomly selected tree trunks in each of the 2500-m² plots. These 40 soil cores from each plot were pooled and subjected to DNA extraction, analysis of soil chemistry, and assignment of climatic variables as described in Tedersoo et al. (2014). For in-depth molecular analysis, we selected 26 samples that represented various ecosystems in PNG and that had relatively well-preserved and inhibitor-free DNA (Suppl. material 1).

To address barcode and primer biases, we selected seven barcodes in SSU (variable domains V4 and V5), ITS (ITS1 and ITS2), and LSU (variable domains D1, D2, and D3) of the nuclear rDNA (Figure 1). These short barcodes represent the most common marker choices in cloning and metabarcoding studies of fungi either alone or combined within a longer stretch of rDNA. To amplify these barcodes, we selected the most commonly used primers for each respective marker, targeting primer pairs that produce amplicons of 250-400 nucleotides on average. For the ITS1 and ITS2 barcodes, we used alternative forward primers to further quantify potential biases arising from primer choice. Due to a lack of suitable primers between the variable D1 and D2 domains of LSU, we designed the LF402 primer and its reverse complement LF402F that are relatively specific to fungi compared to other eukaryotes (Appendix

1). For the V4 domain of SSU, we generated a reverse primer Euk742R that covers a wide range of eukaryotes including the most common fungal groups. In addition to previous modifications to the ITS3 and ITS4 primers (ITS3tagmix and ITS4ngs, respectively; Tedersoo et al. 2014), we also optimized the commonly used SSU515F, SSU1196R, ITS1, ITS1F, and LR0R primers to improve their taxonomic coverage and to balance their melting temperature compared to the respective reverse primers. All primers are detailed in Table 2 and their relative positions are indicated in Figure 1 and on the UNITE homepage (<https://unite.ut.ee/primers.php>). We further evaluated the suitability of these primers through in silico-based assays on SSU and LSU data sets of SILVA release 119 (Quast et al. 2013), the SSU, ITS and LSU data sets of Bellemain et al. (2010), and the ITS reference data set of UNITE (Abarenkov et al. 2010). In addition to manual comparison of primer and template sequences, we used EcoPCR (www.grenoble.prabi.fr/trac/ecoPCR) for SSU and LSU. The mismatched primer variants were further BLASTn-searched against the International Nucleotide Sequence Database consortium (INSDc) to determine the natural frequency of inferred mismatches and to distinguish these from rare sequencing errors. The primer mismatches with fungal templates are outlined in Appendix 1.

A reverse or forward primer for each barcode was supplemented with one of the sixteen 10-base identifier tags (Table 1). We used a mock community sample with known composition of 24 species (Suppl. material 2) and a soil sample from Alaska as positive controls. We also added negative controls and spiked two soil samples (G2655 and G2658) with 2% genomic DNA from *Archaeorhizomyces finlayi* (isolate Ny10) and *A. borealis* (isolate 600) to screen accumulation of their sequences in primer pairs that have no or two mismatches (ITS4ngs and LR0Rngs) to the Archaeorhizomycetes. The PCR cocktail consisted of 0.6 µl DNA extract, 0.5 µl each of the primers (20 pmol), 5 µl 5xHOT FIREPol Blend Master Mix (Solis Biodyne, Tartu, Estonia), and 13.4 µl double-distilled water. PCR was carried out in four replicates using the following thermocycling conditions: an initial 15 min at 95 °C, followed by 30 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, and a final cycle of 10 min at 72 °C. Four replicate PCR products were pooled and their relative quantity was estimated by running 5 µl amplicon DNA on 1% agarose gel for 15 min. DNA samples yielding no visible band were re-amplified using 35 cycles in an effort to obtain sufficient PCR product, whereas samples with a very strong band were re-amplified with only 25 cycles. Amplicons were subjected to quantity normalization with a SequalPrep Normalization Plate Kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. Normalized amplicons were subjected to ligation of Illumina adaptors using two variants of the TruSeq DNA PCR-free HT Sample Prep kit (Illumina Inc., San Diego, CA, USA). All samples were sequenced in a single Illumina MiSeq 2×300 paired-end run. HTS sequences are available in Short Read archive under accession SRP055957.

Selected soil samples (Suppl. material 1) were also subjected to shotgun sequencing of the full metagenome. DNA was fragmented to 300–400 bases on average with the Covaris S2 instrument (Covaris Ltd, Woburn, MA, USA). Samples were prepared for sequencing with the TruSeq Nano DNA HT Library Prep Kit (Illumina Inc.) accor-

Table 1. Primers and identifier tags used for Illumina MiSeq sequencing in this study.

Primer name	Features	Primer sequence	barcode	Reference
SSU515Fngs	Fwd, tagged	GCCAGCAGCCGGCGTAA	SSU V4	This study
Euk742R	Rev	AAATCCAAGAATTCACCTCT	SSU V4	This study
SSU817F	Fwd	TTAGCATGGAAATAATRAATAGGA	SSU V5	Borneman and Hartin 2000
S1196Rngs	Rev, tagged	TCTGGACCTGGTGAGTTT	SSU V5	This study
ITS1Fngs	Fwd, tagged	GGTCATTTAGAGGAAGTAA	ITS1 (combination 1)	This study
ITS1ngs	Fwd, tagged	TCCGTAGGTGAACCTGC	ITS1 (combination 2)	This study
ITS2	Rev	GCTGGCTTCATCGATGC	ITS1 (combinations 1,2)	White et al. 1990
ITS3tagmix1*	Fwd	CTAGACTCGTCATCGATGAAGAACCAG	ITS2 (combination 1)	Tederso et al. 2014
ITS3tagmix2*	Fwd	CTAGACTCGTCAACCGATGAAGAACCAG	ITS2 (combination 1)	Tederso et al. 2014
ITS3tagmix3*	Fwd	CTAGACTCGTCAACCGATGAAGAACCAG	ITS2 (combination 1)	Tederso et al. 2014
ITS3tagmix4*	Fwd	CTAGACTCGTCATCGATGAAGAACCAGTAG	ITS2 (combination 1)	Tederso et al. 2014
ITS3tagmix5*	Fwd	CTAGACTCGTCATCGATGAAGAACCAGTGG	ITS2 (combination 1)	Tederso et al. 2014
gITS7	Fwd	GTGARTCATCGARTCTTTG	ITS2 (combination 2)	Ihrmark et al. 2012
ITS4ngs	Rev, tagged	TTCCCTSCGCTTATTTGATATGC	ITS2 (combinations 1,2)	Tederso et al. 2014
LR0Rngs	Fwd, tagged	ACSGGTGAACCTTAAGC	LSU D1	This study
LF402	Rev	TTCCCTTTYARCAATTTTCAC	LSU D1	This study
LF402Fmix1	Fwd	GTGAAATTYTTRAAAGGGAA	LSU D2	This study
LF402Fmix3	Fwd	GTGAAATYTCAAAAAGGGAA	LSU D2	This study
TW13	Rev, tagged	GGTCCGTGTTTCAAGAGC	LSU D2	T.J. White unpublished
LR3R	Fwd	GTCTTGA AACACGGGACC	LSU D3	Hopple and Vilgalys 1994
LR5	Rev, tagged	TCCTGAGGGAAAACCTTCG	LSU D3	Hopple and Vilgalys 1994
Tag 001	Tag	ACGAGTGCCT	All	Tederso et al. 2014
Tag 002	Tag	ACGCTCGACA	All	Tederso et al. 2014
Tag 003	Tag	AGACGCCACTC	All	Tederso et al. 2014
Tag 026	Tag	ACATACGCCGT	All	Tederso et al. 2014
Tag 028	Tag	ACTACTATGT	All	Tederso et al. 2014
Tag 029	Tag	ACTGTACAGT	All	Tederso et al. 2014

Primer name	Features	Primer sequence	barcode	Reference
Tag 030	Tag	AGACTATACT	All	Tedersoo et al. 2014
Tag 032	Tag	AGTACGCTAT	All	Tedersoo et al. 2014
Tag 033	Tag	ATAGAGTACT	All	Tedersoo et al. 2014
Tag 049	Tag	ACGGGATCGA	All	Tedersoo et al. 2014
Tag 050	Tag	ACTAGCAGTA	All	Tedersoo et al. 2014
Tag 052	Tag	AGTATACATA	All	Tedersoo et al. 2014
Tag 053	Tag	AGTCGAGAGA	All	Tedersoo et al. 2014
Tag 054	Tag	AGTGCTACGA	All	Tedersoo et al. 2014
Tag 077	Tag	ACGACAGCTC	All	Tedersoo et al. 2014
Tag 078	Tag	ACGTCTCATC	All	Tedersoo et al. 2014

* The first 10 bases (CTAGACTCGT) represent an inert tag that does not align to any organism.

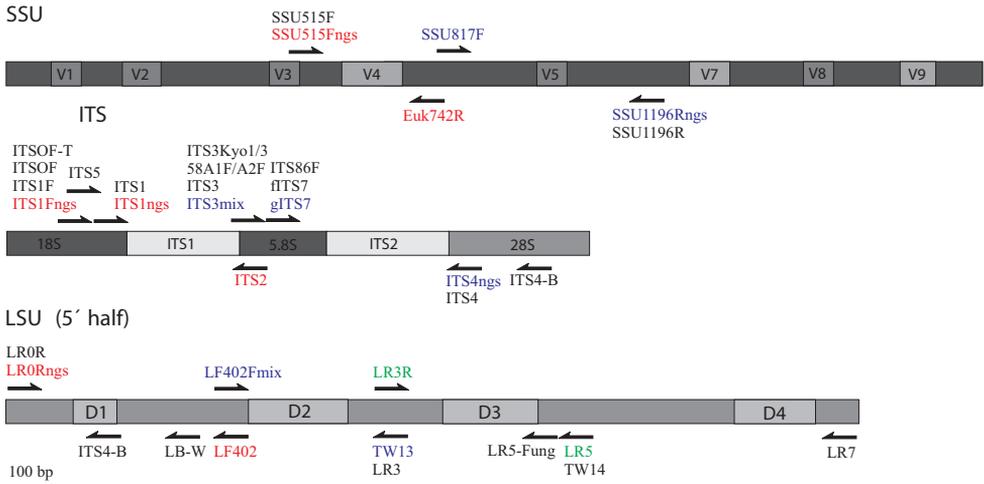


Figure 1. Map of ribosomal DNA indicating variable regions as well as primers used and/or discussed in this study. Primers pairs used for HTS are highlighted.

Table 2. Number of sequences recovered using different barcode-primer pair combinations.

Primer pair	Raw sequences	Quality-filtered sequences	Fungal sequences	% fungal sequences	Data set connectance
SSU515Fngs-Euk742R	2156146	1751042	1177111	67.2	0.264
SSU817F-SSU1196Rngs	1583096	1431850	1382433	96.5	0.340
ITS1Fngs-ITS2	1104540	697900	634098	90.9	0.128
ITS1ngs-ITS2	1025094	451500	327397	72.5	0.128
ITS3tagmix-ITS4ngs	2665289	1943355	1706010	87.8	0.065
gITS7-ITS4ngs	1293599	1005751	923170	91.8	0.062
LR0Rngs-LF402	1001017	743637	742973	99.9	0.201
LF402Fmix-TW13	101161	84282	64661	76.7	nd
LR3R-LR5	761164	567222	384357	67.8	0.359

ding to the instructions of the manufacturer. These samples were sequenced using the Illumina HiSeq 2×250 paired-end protocol covering 80% of a full plate.

We amplified DNA from two soil samples (G2655 and G2658 that were spiked with *A. finlayi* and *A. borealis* DNA) using the ITSOF and LR7 primers (Hopple and Vilgalys 1994; Tedersoo et al. 2008) and *Pfu* DNA Polymerase (Fermentas, Kaunas, Lithuania) to produce 2000–2200 base-pair (bp) amplicons for improved classification. The PCR mix included 2.5 µl 10× *Pfu* buffer, 0.4 µl *Pfu* polymerase, 0.25 µl dNTP mix (20 mM each), 0.7 µl each primer (20 µmol), 2 µl template DNA, and 18.45 µl ddH₂O. Thermocycling conditions included an initial denaturation at 95 °C for 3 minutes, 30 cycles of denaturation for 30 seconds, annealing at 53 °C for 30 seconds, extension at 72 °C for 5 minutes, and a final extension for 10 minutes. The amplicons were cloned using StrataClone Blunt PCR Cloning kit (Agilent Technologies, La Jolla, CA, USA) following the manufacturer's instructions. Positive inserts were re-amplified us-

ing primers ITSOF and LR5 (Hopple and Vilgalys 1994). Amplicons were sequenced bidirectionally using primers ITS5 (White et al. 1990) and LR5. Unique sequences are available in the UNITE database (accessions UDB014801-UDB014812).

Bioinformatics

Paired-end sequencing (2×300 bp) in the Illumina MiSeq sequencer resulted in 12,771,565 reads. LSU and SSU amplicons were paired, quality filtered, and demultiplexed using the LOTUS pipeline (Hildebrand et al. 2014) with the following options: overlap >9 bases, average quality >28, no DNA ambiguities allowed, no differences to primer+identifier allowed, homopolymer length <9, and minimum sequence length >199 bases. Primers and tags were removed and reads were trimmed to 200 bases covering the variable domain, and clustered with cd-hit 4.6.1 (Fu et al. 2012) at 97% sequence similarity into operational taxonomic units (OTUs). Chimeric OTUs were detected using the UCHIME de novo and reference-based algorithms as implemented in USEARCH 7 (Edgar et al. 2011). The SILVA 119 data set was used as reference corpus.

The ITS reads were quality filtered using MOTHUR 1.33.3 (Schloss et al. 2009) with the following options: average quality over 15 nucleotide window >30, no ambiguities allowed, and homopolymer length unrestricted. The remaining reads were paired using the PANDASEQ assembler (Masella et al. 2012) with minimum overlap of 25 bases and demultiplexed using mothur (no differences to primer+identifier allowed, minimum sequence length >200 bases). ITS1 and ITS2 sequences were processed in ITSx 1.0.9 (Bengtsson-Palme et al. 2013) to remove flanking 18S, 5.8S and 28S rRNA gene fragments. Potential chimeras were removed as described above, using the UNITE-UCHIME release as reference (Nilsson et al. in press). The few reads containing primer sequences were excluded using Unix commands. The quality-filtered ITS1 and ITS2 sequences were clustered as described above.

Following exclusion of singletons from all HTS data sets (cf. Brown et al. 2015), representative sequences were selected for each OTU based on the greatest similarity to the consensus sequence. These sequences were subjected to taxonomic assignments using the NAÏVE BAYESIAN CLASSIFIER (NBC; Wang et al. 2007) and BLASTn (options: word size=7, gap penalty=-1, gap extension penalty=-2 and match score=1) searches retrieving 10 best BLASTn matches. For both methods, The SILVA 119 eukaryote data set and the INSDc fungal data set served as references for SSU and LSU. The UNITE 7.0beta data set (<https://unite.ut.ee/repository.php>) and the INSDc eukaryote data set were used as a reference for BLASTn- and NBC-based taxonomic assignment of ITS, respectively.

Because of differential taxonomic resolution among the barcodes, we used the taxonomic assignments of both NBC and BLASTn searches to complement each other as both methods alone provided no assignment for ca 40% of OTUs due to poor representation of fungal data in SILVA, obvious misidentifications in INSDc, and great abundance of taxonomically unassigned sequence data (resulting in poor reso-

lution using NBC). To optimize classification, we therefore combined and verified results of different methods and determined approximate E -value and sequence similarity thresholds for robust identification to phylum or class level for each barcode. For the SSU barcodes, we determined that sequences with a BLASTn E -value $<e^{-80}$ and $>90\%$ similarity to verified fungal sequences could be reliably assigned to the fungal kingdom. Sequence similarities at or greater than 95% served to provide class-level assignments. Taxa with short introns in the SSU V5 barcode (primers pair SSU817F-SSU1196Rngs) such as some members of Pezizomycotina and Agaricomycetes exhibited more variation, and these groups were therefore manually evaluated for class-level affinity. For ITS1, ITS2, and LSU D2 (primers LF402Fmix-TW13), we determined that BLASTn E -values $<e^{-50}$ and sequence similarity $>75\%$ over $>70\%$ sequence length allowed robust assignment to the fungal kingdom. For individual classes, we relied on an 80% sequence similarity threshold, except the early diverging lineages, Archaeorhizomycetes, and Cantharellales, where we used 75% sequence similarity. For the LSU D1 barcode (LR0Rngs-LF402 primers), BLASTn E -values $<e^{-80}$ and sequence similarity above 80% were sufficient to consider sequences to be fungal, whereas similarity $>85\%$ was indicative of class-level affiliation. For the D3 barcode of LSU (LR3R-LR5 primer pair), BLASTn E -value $<e^{-100}$ and sequence similarity $>85\%$ to fungi was strongly suggestive of fungal origin, and sequence similarity $>90\%$ allowed placement to classes. Many Leotiomycetes and Eurotiomycetes possessed introns close to the LR5 primer site, which allowed identification of these groups at $>80\%$ similarity level. For all barcodes, phylogenetic placement of BLASTn matches and NBC were highly concordant, except several instances in Glomeromycota, Microbotryomycetes, and Tremellomycetes. For these groups, we relied on the 10 best BLASTn matches.

We followed the taxonomy of INSDc, except raising several early diverging lineages to phylum rank (cf. Tedersoo et al. 2014). The OTUs of ITS1 and ITS2 barcodes were assigned to EcM lineages following Tedersoo and Smith (2013). For several common lineages such as /hebeloma-alnicola, /cortinarius, /cenococcum and /meliniomyces, SSU and LSU barcodes provided insufficient discrimination from non-EcM relatives such that it was impossible to reliably assign ecological categories to SSU and LSU data.

For the shotgun metagenome data, samples were demultiplexed, and LSU and SSU regions of all organisms were extracted using SORTMERA (Kopylova et al. 2012). Ribosomal DNA sequences were classified against the SILVA 119 reference database (SSU, ITS, and LSU) and UNITE 7.0beta (ITS region). We also generated sets of 10 best BLASTn matches against the INSDc fungal data set. We manually checked conflicting assignments as well as those with $<95\%$ similarity to fungal sequences to improve phylum- and class-level classification.

To understand potential amplification biases related to sequence length in the ITS1 and ITS2 barcodes, we downloaded all ITS sequences of the 16 most common fungal classes (based on our amplicon data) from UNITE 7.0beta data set. Ribosomal RNA genes flanking the ITS1 and ITS2 barcodes were trimmed using ITSx 1.0.9. Average and median values and standard deviations were calculated for each group.

Statistical analyses

For OTU-based statistical analyses, we removed all non-fungal sequences and rarefied all amplicon samples to a depth of 8609 sequences using MOTHUR. This depth represents the median number of sequences of the ITS1 (ITS1ngs-ITS2 primers) barcode that was the second lowest among all markers (Table 2). The LSU D2 barcode comprised several times less sequences, and this marker was therefore removed from comparisons of taxonomic richness and statistical performance of barcodes, but kept for comparisons of taxonomic coverage. Soil sample G2653 was poorly covered by sequence data for three barcodes (<1000 sequences) and was therefore omitted from all analyses. Samples G2666 (SSU V5 barcode), G2677 (SSU V4 barcode), and G2678 (LSU D3 barcode) were variously poorly represented in single data sets. Their values were treated as missing data in the statistical analyses.

OTU accumulation curves and their 95% confidence intervals were computed for ITS1 and ITS2 barcodes using ESTIMATES 9.1.0 (Colwell 2013). To further address whether ITS1Fngs and ITS1ngs or ITS3tagmix and gITS7 primers exhibit additive performance (i.e., recover more OTUs when combined), we pooled the primer pair data by samples and rarefied these combined samples to 8609 sequences, followed by inference of OTU accumulation curves as described above.

Differences in OTU richness among samples and barcode-primer combinations were evaluated based on two-way main-effect ANOVAs supplemented with Unequal n HSD tests for multi-level comparisons. To address the relative performance of the eight barcode-primer pair combinations in recovering the role of spatial, edaphic, floristic, and climatic predictors on fungal community composition, we performed multivariate permutational ANOVAs as implemented in the ADONIS routine of the *vegan* package of R (R Core Development Team 2013). Geographical coordinates were translated into Principal Coordinates of Neighbouring Matrices (PCNM) vectors with soil element concentrations being logarithm-transformed prior to analyses. All four categories of variables were included in separate matrices for these analyses following Pólme et al. (2014). We used both Bray-Curtis dissimilarity and Hellinger distance in analyses of community composition. We also analyzed the relative importance of primer choice for the ITS1 and ITS2 barcodes in ADONIS. Because data sets differed substantially in matrix connectance, we tested whether this matrix property affects the adjusted determination coefficients (R^2_{adj}) that are widely used to interpret model performance quantitatively through the explained variance. Therefore, we re-analyzed the ITS1Fngs-ITS2 data set at 12 levels of connectance by selectively removing species based on their frequency from one (4%) to 12 (48%) and calculated the correlation between the logarithm of connectance and R^2_{adj} as revealed from ADONIS analyses.

To further test changes in phylogenetic community structure among samples and barcode-primer combinations, we assigned the OTUs to fungal classes. Sequence number-based proportions of classes were log-ratio transformed in relation to the proportion of non-Agaricomycetes and unassigned fungi to ensure statistical independence of these groups. The proportion of EcM and arbuscular mycorrhizal (AM) to

non-mycorrhizal sequences was similarly log-ratio transformed but only analyzed for the four ITS data sets. Using the Hellinger distance and Bray-Curtis distance, the effects of barcode-primer pair combination were analyzed using ADONIS. Global Non-metric Multidimensional Scaling (GNMDS) graphs were drawn in parallel using the same options. 95% confidence ellipses were calculated using ORDISURF routine in the Vegan package of R.

Potential analytical biases in the recovery of fungal classes by ribosomal DNA region (SSU, ITS, and LSU) and analysis method (metagenomics and amplicon) were addressed based on the average values of 14 shared samples using two-way main-effect ANOVAs neglecting interactions. rDNA region-based biases in amplicon and metagenomics data sets were further tested using two-way ANOVAs including the samples and regions as fixed factors.

To understand possible amplification biases related to sequence length in the ITS1 and ITS2 barcodes, we downloaded all ITS sequences of the 16 most common fungal classes from UNITE 7.0beta. Ribosomal RNA genes flanking the ITS1 and ITS2 barcodes were trimmed using ITSx 1.0.9. Average and median values and standard deviations were calculated for each group for illustrative purpose.

Results

Amplicons

Although combinations of samples and primer pairs were normalized separately, sequences assigned to each primer pair were differentially represented in the raw and final data sets (Table 2). Much of the data loss from a total of 11,691,106 demultiplexed sequences was attributable to unpaired reads and stacked reads, where certain reads consisted of two repeats of itself. Some of these incomplete repeats could not be successfully filtered out by bioinformatics tools, but these sequences were easily recognized due to their abnormal length (>480 bases) and were removed manually. Remarkably, one such artefactual OTUs formed the second most abundant taxon in the ITS2 data set. The unequal occurrence of barcode-marker combinations was not related to the average marker length as the least represented LSU D2 (LF402Fmix-TW13) barcode was the shortest among all barcodes. Therefore, we recommend not to multiplex different barcodes in the same libraries unless their roughly equal representation is known.

Barcodes generated by the universal primer pairs SSU515Fngs-Euk742R (SSU V4) and LR3R-LR5 (LSU D3) exhibited the distinctly lowest proportion of fungal sequences (67–68%), suggesting that fungi account for roughly two thirds of eukaryote ribosomal DNA in the studied soils on average. The proportion of fungal sequences was the greatest for the primer pair LR0Rngs-LF402, reaching 99.9% of all sequences. Classifications based on both BLASTn searches and NBC individually assigned >90% of the reads to fungi, indicating that this primer pair could indeed be the most fungus-specific of those tested. Of fungal classes, Agaricomycetes was the dominant group in all data sets.

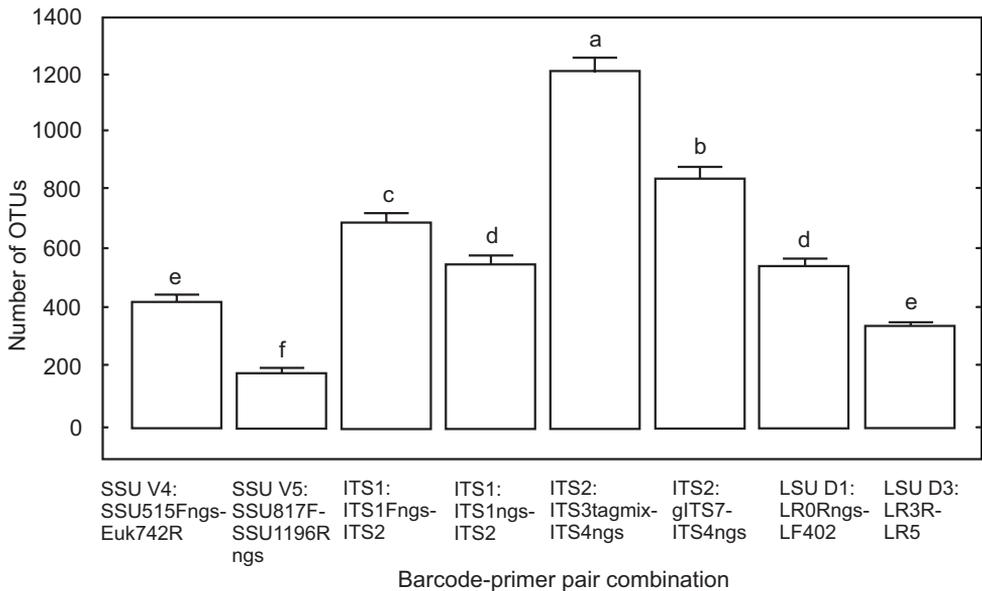


Figure 2. Sample-based OTU richness as recovered by different barcode-primer pair combinations. Error bars denote standard error; different letters indicate statistically different groups.

The barcode-primer combinations exhibited five-fold differences in the number of fungal OTUs recovered in total and on the basis of samples rarefied to 8906 sequences (Figure 2; Suppl. material 3–6). As expected, the most conserved barcode - SSU V5 (SSU817F-SSU1196Rngs) - revealed the lowest number of OTUs, followed by LSU D3 (LR3R-LR5) and SSU V4 (S515Fngs-Euk742R) barcodes. Of rDNA genes, the LSU D1 barcode recovered greater OTU richness compared to other SSU and LSU barcodes. Overall, however, the ITS1 and ITS2 barcodes recovered the greatest number of OTUs. Both primer pairs targeting ITS2 sequences produced significantly more OTUs than those used to amplify ITS1, although the latter is considered somewhat less conserved (Nilsson et al. 2009; Wang et al. 2015). For the ITS2 barcode, the forward primer ITS3tagmix produced significantly more OTUs per sample compared with the gITS7 primer (especially Leotiomycetes and Lecanoromycetes), but their richness estimates were strongly correlated ($n=25$; $R=0.96$; $P<0.001$; Table 3). Similarly, of the two primers targeting the ITS1 barcode, the ITS1Fngs forward primer produced significantly more OTUs compared to the ITS1ngs primer (especially Sordariomycetes), but again these estimates were strongly correlated ($n=25$; $R=0.83$; $P<0.001$). OTU accumulation curves of separate and mixed ITS1 data sets indicated that combining the results of both primer pairs enabled recovery of 6.6% more OTUs, but the fraction of gITS7 data did not add to the ITS3tagmix data (Figure 3). Overall, primer pairs for the ITS1 barcode were the least correlated with other barcodes in predicting OTU richness and the median value of richness. The LSU D1 barcode and ITS2 barcode (ITS3tagmix-ITS4ngs) had the strongest average correlation with other combinations.

Table 3. Correlation matrix of eight barcode-primer combinations in recovering OTUs per sample rarefied to 8609 sequences. Values denote Pearson correlation coefficients. Values <0.44 are statistically not significant at 95% confidence level.

	Median	SSU515Fngs-Euk742R	SSU817F-SSU1196Rngs	ITS1Fngs-ITS2	ITS3tagmix-ITS4ngs	gITS7-ITS4ngs	LR0Rngs-LF402
SSU515Fngs-Euk742R	0.82						
SSU817F-SSU1196Rngs	0.81	0.88					
ITS1Fngs-ITS2	0.75	0.41	0.43				
ITS1ngs-ITS2	0.81	0.41	0.49	0.83			
ITS3tagmix-ITS4ngs	0.85	0.71	0.69	0.69	0.67		
gITS7-ITS4ngs	0.82	0.66	0.69	0.70	0.63	0.96	
LR0Rngs-LF402	0.90	0.93	0.87	0.48	0.50	0.81	0.77
LR3R-LR5	0.82	0.84	0.81	0.44	0.55	0.71	0.64
							0.85

Table 4. Pearson correlations among barcode-primer pair combinations and the soil metagenome in recovering the relative abundance of fungal classes. All correlations are statistically highly significant ($P < 0.001$).

	SSU515Fngs-Euk742r	SSU817F-SSU1196Rngs	ITS1Fngs-ITS2	ITS1ngs-ITS2	ITS3tagmix-ITS4ngs	gITS7-ITS4ngs	LR0Rngs-LF402	LF402F-TW13	LR3R-LR5	Median amplicon
SSU817F-SSU1196Rngs	0.82									
ITS1Fngs-ITS2	0.89	0.80								
ITS1Fngs-ITS2	0.89	0.74	0.97							
ITS3tagmix-ITS4ngs	0.88	0.76	0.92	0.90						
gITS7-ITS4ngs	0.87	0.74	0.92	0.91	0.98					
LR0Rngs-LF402	0.87	0.75	0.90	0.89	0.91	0.90				
LF402F-TW13	0.87	0.72	0.88	0.88	0.85	0.84	0.89			
LR3R-LR5	0.82	0.79	0.79	0.75	0.72	0.71	0.79	0.79		
Median amplicon	0.92	0.80	0.96	0.95	0.98	0.97	0.95	0.91	0.79	
Metagenome	0.83	0.77	0.85	0.84	0.82	0.80	0.86	0.85	0.76	0.87

Across all samples, barcode-primer pair correlations were strongly correlated in recovering the relative abundance of fungal classes (Table 4). These correlations were slightly stronger for primer pairs targeting the same barcode. The ITS3tagmix-ITS4ngs and gITS7-ITS4ngs primer pairs exhibited the strongest correlations to the median values across all barcodes. On average, the most conserved LSU D3 (LR3R-LR5) and SSU V5 (SSU817F-SSU1196Rngs) barcodes had the distinctly lowest correlations to other barcodes and the median values of all barcodes taken together.

Ecological analyses using all barcodes consistently revealed that vegetation structure was the strongest predictor of fungal communities (Table 5). Contrary to our predictions, the most conservative barcodes SSU V5 (SSU817F-SSU1196Rngs) and LSU D3 (LR3R-LR5) indicated that floristic variables explain the greatest amount of variation in the fungal community among all barcodes. Conversely, floristic variables accounted for the lowest amount of variation among all barcodes in all four ITS data sets. We suspected that lower matrix connectance could result in relatively lower detection power in the more sensitive and larger data sets. Successive removal of infrequent species indicated that the logarithm of connectance was extremely strongly correlated with the coefficient of determination up to connectance values of 0.45 (Bray-Curtis distance: $n=8$; $R=0.980$; $P<0.001$; Hellinger distance: $n=8$; $R=0.995$; $P<0.001$). Beyond this point, the adjusted determination coefficients reached a plateau or began to decline (Figure 4).

In the ITS1 barcode data set, the choice of primers (ITS1Fngs vs. ITS1ngs) explained 2.0% and 1.8% of the community variation based on Hellinger distance (partial analysis: $F_{1,33}=3.12$; $P<0.001$) and Bray-Curtis dissimilarity ($F_{1,33}=3.12$; $P<0.001$), respectively. This amount of variation was marginal compared to the effects of vegetation (16.0–20.0%) and edaphic parameters (5.5–6.4%) but comparable to climatic effects (2.8–3.8%). In the ITS2 barcode data set, the primer pair (ITS3tagmix vs gITS7) accounted for <1% variation in fungal community composition based on both distance measures ($P<0.001$); this was negligible compared to the effects of vegetation (19.6–22.2%), spatial vectors (5.5–6.9%), soil variables (5.6–5.8%), and climatic predictors (3.3–3.4%). The class-level analysis of all nine barcode-primer pair combinations revealed that primer pair was the strongest overall predictor, explaining 38.1% of the community variation (partial analysis: $F_{8,182}=38.3$; $P<0.001$; Figure 5), overshadowing the environmental effects (altogether 22.7%).

Metagenomics

Of the 290,779,313 high-quality metagenome sequences from the PNG soil samples, 1,309,342 (0.45%) were assigned to ribosomal DNA of prokaryotic and eukaryotic organisms. Bacterial sequences and eukaryote sequences unassigned to any kingdom dominated the rDNA subset. Only 16,833 (1.29%) of these sequences were determined to represent fungal nuclear rDNA. Across all samples and regions, Agaricomycetes (41.4%), Eurotiomycetes (10.4%), and Dothideomycetes (5.5%) dominated,

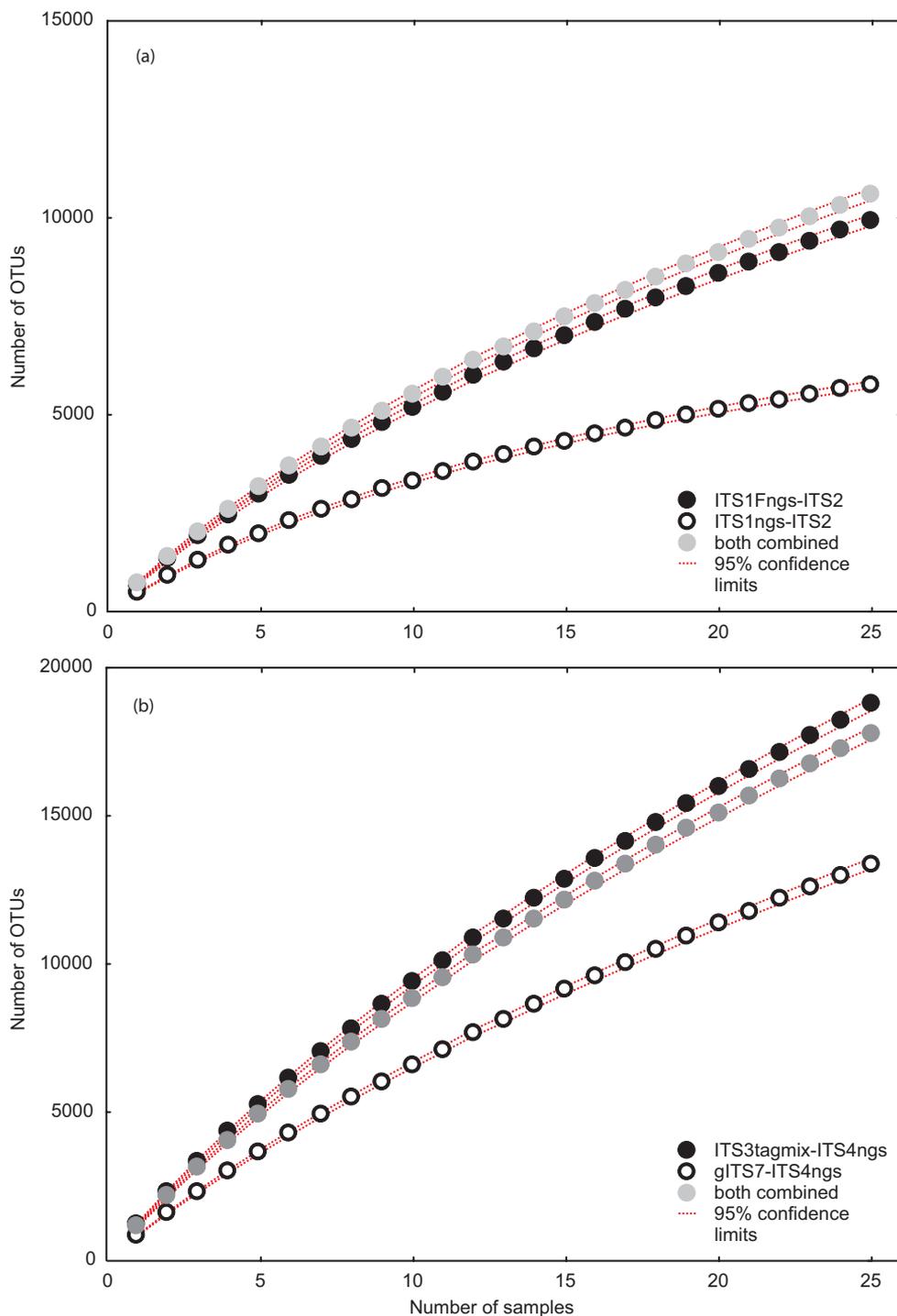


Figure 3. Rarefied OTU accumulation curves for samples based on the (a) ITS1 and (b) ITS2 barcodes and their 95% confidence intervals.

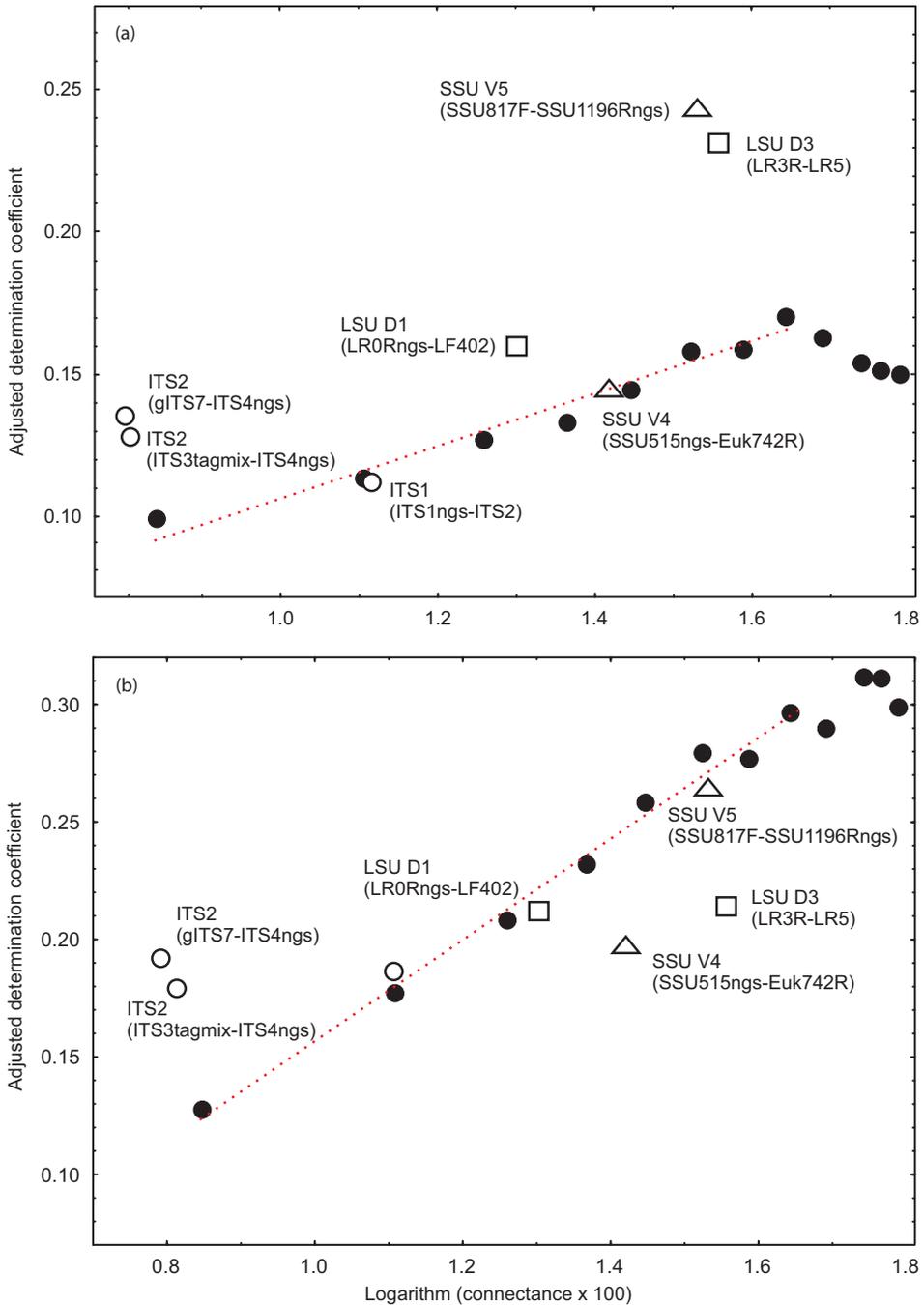


Figure 4. Relationship between connectance and adjusted coefficient of determination (R^2_{adj}) for floristic variables across different barcode-primer pair combinations based on (a) Bray-Curtis distance and (b) Hellinger distance. Pointed line indicates correlation in the ITS1FnGs-ITS2 data set (filled circles), covering eight connectance classes ($C < 0.45$). Open circles, other ITS1 and ITS2 primer pairs; triangles, SSU barcodes; rectangles, LSU barcodes.

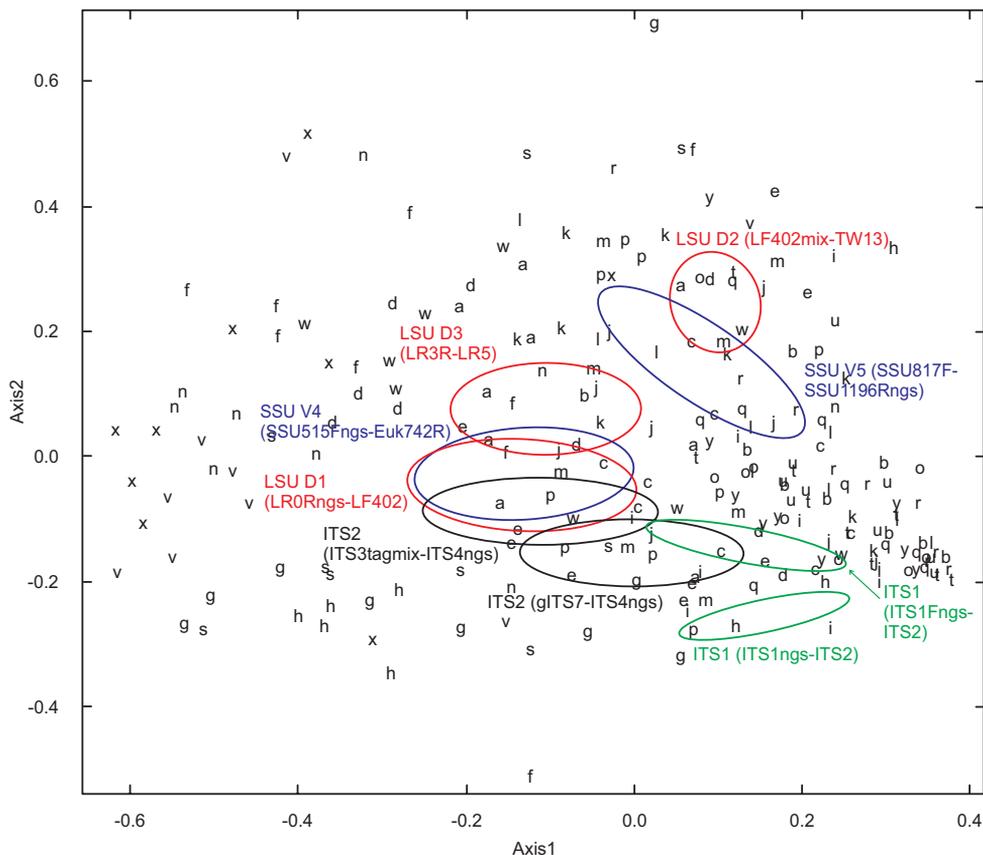


Figure 5. Global Nonmetric Multidimensional Scaling (NMDS) graph demonstrating the relative placement of samples (lower case letters, encoded in Suppl. material 1) in the ordination space. 95% confidence ellipses are indicated for each barcode-primer pair combination. For two-dimensional solution, stress=0.191 ($R^2=0.875$).

whereas 13.2% sequences could not be assigned to any fungal class (Suppl. material 7). The fungal class Mixiomycetes was only recovered based on metagenome analysis, but many other uncommon groups in the amplicon data set were absent from the metagenome data set, presumably due to the comparatively low number of fungal metagenome sequences. The relative abundance of fungal classes in the amplicon-free soil metagenome was significantly correlated with that of all barcodes; the two most conserved barcode-primer pair combinations exhibited the weakest correlations (Table 4).

Within the soil metagenome, there were substantial differences in the recovery of fungal classes based on SSU, ITS, and LSU (Figure 6). Both in the metagenome and amplicon data set, 11 out of the 16 most common fungal classes exhibited significant differences in recovery by the rDNA markers. In spite of PCR and primer biases, the taxonomic composition of fungi was more similar based on the same marker using

Table 5. Effects of environmental parameters on community composition of fungi as revealed by eight rDNA barcode-primer pair combinations and two distance measures.

	Bray-Curtis dissimilarity					Hellinger distance				
	DF	Sum of Squares	F-value	R^2_{adj}	P-value	DF	Sum of Squares	F-value	R^2_{adj}	P-value
SSU515Fngs-Euk742R										
Vegetation	6	2.401	1.659	0.145	0.003	6	2.049	2.131	0.197	0.001
Climate	3	1.011	1.397	0.028	0.079	3	0.592	1.231	-0.015	0.127
Soil	5	1.071	0.888	-0.068	0.714	5	0.995	1.241	-0.026	0.075
Spatial vectors	1	0.104	0.430	-0.029	0.988	1	0.122	0.761	-0.020	0.811
Residuals	8	1.930	0.296			8	1.282	0.254		
SSU817F-SSU1196Rngs										
Vegetation	6	2.139	2.161	0.245	0.001	6	1.369	2.666	0.264	0.001
Climate	3	0.737	1.490	0.025	0.115	3	0.355	1.383	-0.014	0.085
Soil	5	0.605	0.733	-0.118	0.837	5	0.507	1.184	-0.062	0.212
Spatial vectors	1	0.039	0.234	-0.037	0.991	0	n.a.	n.a.	n.a.	n.a.
Residuals	8	1.320	0.273			9	0.771	0.257		
ITS1Fngs-ITS2										
Vegetation	6	3.682	1.596	0.113	0.002	6	3.796	2.087	0.177	0.001
Soil	5	2.139	1.113	-0.017	0.146	5	1.870	1.234	-0.025	0.038
Climate	3	1.283	1.112	-0.009	0.201	3	1.155	1.269	-0.010	0.055
Spatial vectors	2	0.810	1.053	-0.011	0.349	2	0.670	1.105	-0.017	0.247
Residuals	8	3.153	0.285			8	2.425	0.245		
ITS1ngs-ITS2										
Vegetation	6	3.682	1.549	0.113	0.001	6	3.998	2.069	0.187	0.001
Soil	5	2.139	1.080	-0.017	0.241	5	1.919	1.192	-0.027	0.086
Climate	3	1.283	1.079	-0.009	0.286	3	1.166	1.207	-0.013	0.104
Spatial vectors	3	1.112	0.935	-0.027	0.731	3	0.917	0.949	-0.041	0.612
Residuals	7	2.773	0.252			7	2.254	0.220		
ITS3tagmix-ITS4ngs										
Vegetation	6	3.7079	1.723	0.129	0.001	6	3.743	2.069	0.179	0.001
Soil	5	2.0253	1.129	-0.024	0.102	5	1.820	1.207	-0.027	0.058
Climate	3	1.3506	1.255	0.001	0.023	3	1.166	1.289	-0.006	0.027
Spatial vectors	3	1.1044	1.026	-0.025	0.398	3	0.895	0.989	-0.038	0.498
Residuals	7	2.5107	0.235			7	2.111	0.217		
gITS7-ITS4ngs										
Vegetation	6	3.806	1.726	0.136	0.001	6	3.921	2.178	0.192	0.001
Soil	5	2.0525	1.117	-0.023	0.154	5	1.864	1.243	-0.027	0.037
Climate	3	1.3146	1.193	-0.004	0.088	3	1.185	1.316	-0.007	0.037
Spatial vectors	1	0.3236	0.881	-0.012	0.721	1	0.282	0.940	-0.014	0.545
Residuals	9	3.3071	0.306			9	2.701	0.271		
LR0Rngs-LF402										
Vegetation	6	3.254	1.897	0.160	0.001	6	2.869	2.259	0.212	0.001
Soil	5	1.792	1.253	-0.006	0.049	5	1.357	1.282	-0.019	0.034
Climate	3	0.983	1.146	-0.015	0.198	3	0.730	1.150	-0.024	0.208

	Bray-Curtis dissimilarity					Hellinger distance				
	DF	Sum of Squares	F-value	R^2_{adj}	P-value	DF	Sum of Squares	F-value	R^2_{adj}	P-value
Spatial vectors	3	0.768	0.895	-0.043	0.741	3	0.578	0.911	-0.049	0.704
Residuals	7	2.001	0.227			7	1.482	0.211		
LR3R-LR5										
Vegetation	6	2.328	2.309	0.231	0.001	6	1.408	2.339	0.214	0.001
Soil	5	0.821	0.977	-0.083	0.529	5	0.618	1.232	-0.043	0.1
Climate	3	0.582	1.154	-0.026	0.248	3	0.359	1.194	-0.027	0.171
Spatial vectors	1	0.316	1.878	0.016	0.035	1	0.175	1.743	0.009	0.027
Residuals	8	1.345	0.249			8	0.802	0.239		

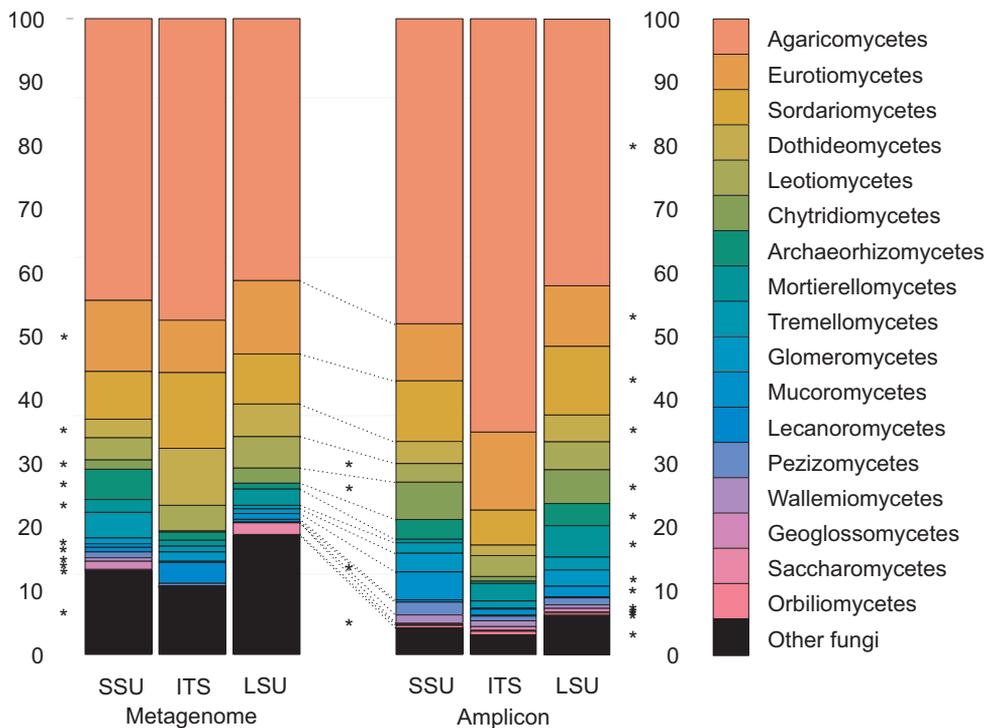


Figure 6. Relative abundance of fungal classes in the amplicon and metagenomics data sets divided into SSU, ITS, and LSU subsets averaged over different barcodes (amplicon data) and 14 shared samples. Asterisks in the margins indicate significant differences in recovery of classes among SSU, ITS, and LSU of metagenomics (right) and amplicon (left) data sets. Asterisks in the center indicate significant differences between the metagenomics and amplicon-bases approaches.

either metagenomics- or amplicon-based HTS. Chytridiomycota and Pezizomycetes were significantly less common in the metagenome compared to the amplicon-based HTS data, but Leotiomycetes and the unassigned category were more abundant in the metagenomics data set (Figure 6).

Metagenomic analysis of the mock community enabled us to estimate class-level biases in identification of fungi. While other groups were roughly evenly represented, *Archaeorhizomyces* spp. accounted for 17.0%, 5.1%, and 10.1% of metagenomics sequences in the SSU, ITS, and LSU data subsets, respectively, which is consistent with trends in the metagenomics and amplicon data sets of soil samples. Across all three markers, 6.2% and 2.4% of sequences respectively belonged to the unknown category or were affiliated with classes that were not included in the mock community (Suppl. material 7). The proportion of unknowns and incorrectly identified sequences were 4.3 and 2.5 times greater for the LSU compared with ITS data subset. The proportion of incorrect identifications is probably much greater, because our mock community represents sequences from the most common classes and misidentifications among these classes cannot be inferred. Therefore, use of positive controls representing a mock community and single isolates is warranted for evaluation of identification biases in metagenomics data sets.

In silico analyses

In-depth analysis of primer mismatches to fungal templates revealed potential systematic biases inherent to different primer pairs (Appendix 1). The fungus-specific ITS1F primer family has central mismatches to multiple fungal groups as well as 3' terminal mismatches to Chytridiomycota, Saccharomycetes, and several taxa within Dothideomycetes. The “universal” ITS1 primer has two central mismatches to most Sordariomycetes. These two primers complement each other very well in terms of taxon coverage and avoidance of introns, except for the fact that both miss most Microsporidia. The ITS2, ITS3, and gITS7 primers exhibit single mismatches to multiple fungal groups especially within the early diverging lineages. The use of a mixture of multiple modified ITS3 variants accounts for the vast majority of these mismatches (Appendix 1; Tedersoo et al. 2014). Due to extreme variation in the 5.8S rDNA, many lineages in the Tulasnellaceae family are missed by all these primers. The same thing goes for the Microsporidia, whose unusual configuration of the ribosomal genes makes this a very challenging group to target in molecular ecology efforts. The ITS4 and LR0R primer families possess two mismatches to Archaeorhizomycetes and one mismatch to most Chaetothyriales, the latter of which was accounted for by a degenerate position in the ITS4ngs and LR0Rngs primers. Both primers are ill-suited for amplification of a large proportion of Tulasnellaceae and Microsporidia due to multiple central mismatches (Taylor and McCormick 2008; Oja et al. 2015). The LF402 and LF402F primers discriminate among most non-fungal groups and cover most of the fungal kingdom, except for a number of taxa within *Cantharellus*, Tulasnellaceae, Ustilaginaceae, *Candida*, and some early diverging lineages. These can be accounted for by adding degenerate positions or constructing primer mixes (Appendix 1). The LR3R, TW13, and LR5 primers are truly universal, with perfect match to nearly all eukaryotes. Some Saccharomycetes taxa exhibit a single central mismatch to all these

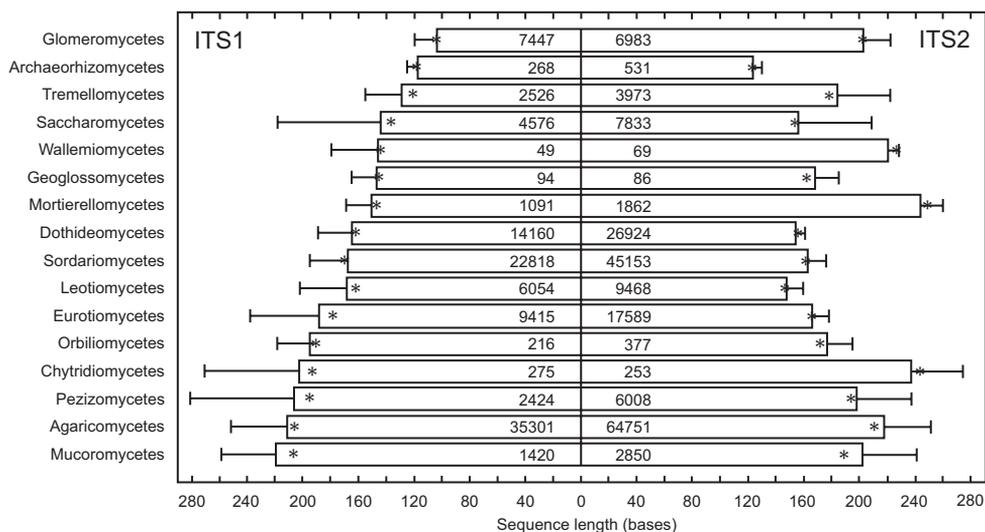


Figure 7. Differences in sequence length in the ITS1 and ITS2 barcodes of 16 most abundant fungal classes as revealed based on amplicon libraries in this study. Columns, asterisks, and error bars represent mean and median values and standard deviation, respectively. Numbers inside bars indicate the number of sequences analyzed (n). Taxa are ordered by average length of the ITS1 region.

primers. The SSU primer SSU515F is nearly universal across the three domains of life. However, Microsporidia and Archaeorhizomycetes exhibit a single central mismatch to this primer. This mismatch can be accounted for by a degenerate position in the original variant of this primer (Turner et al. 1999). The original primer contains an intron site common to many lineages in Saccharomycetes and Pezizomycotina. Shortening the SSU515Fngs primer served to remove the intron position from the priming site in this study. The Euk742R and SSU817F primers possess multiple mismatches to several fungal groups, especially *Cantharellus*, Tulasnellaceae, and Saccharomycetes but also to certain other groups as well. Therefore, these primers are not recommended for fungal metabarcoding. The SSU1196R primer has multiple (including terminal) mismatches to *Cantharellus*, Tulasnellaceae, Saccharomycetes, some groups of Pezizomycetes, and Glomeromycota. Shortening the primer as in SSU1196Rngs ameliorated the mismatches only in part. Of other commonly used primers, the ITS5 primer is universal to most eukaryotes with a single central mismatch to some fungal and non-fungal groups. The LR5F primer has a perfect match to nearly all fungal groups, animals, and stramenopiles, but at the same time it exhibits a terminal mismatch to plants. Of basidiomycete-specific primers, the ITS4-B has multiple mismatches to nearly half of the major orders, but it effectively excludes other fungal phyla and non-fungal groups. The LB-W primer covers nearly all Basidiomycota (except *Cantharellus* and Ustilaginomycetes) and early diverging lineages, but it discriminates strongly against Ascomycota.

Fungal taxa differed roughly three-fold in the length of the ITS1 and ITS2 barcodes (Figure 7). The Archaeorhizomycetes exhibited both very short ITS1 and ITS2

barcodes, with the full ITS region averaging 370 bp. On average, Mortierellomycetes, Glomeromycetes, Tremellomycetes, and Wallemiomycetes had a disproportionately short ITS1 barcode compared with the ITS2 barcode. Except for Wallemiomycetes, the abundance of these groups was lower based on the ITS1 barcode in spite of smaller amplicon size. However, critical mismatches to the ITS1Fngs and ITS2 primers (Appendix 1) may have caused relatively lower recovery of Glomeromycota and Tremellomycetes based on the ITS1.

Discussion

Multiple DNA barcodes

Our analyses of seven barcodes indicate that markers differ substantially in their ability to recover OTUs at the 97% sequence similarity threshold, a threshold value that is almost universally used in HTS studies. Consistent with the lower species-level discrimination power of SSU and LSU compared with ITS (Schoch et al. 2012), both SSU and LSU barcodes recovered fewer OTUs compared to ITS1 and especially ITS2 barcodes. Because the 97% sequence similarity threshold is considered too low for separation of closely related species for the ITS region (Kóljalg et al. 2013), the true taxonomic richness is probably underestimated with all barcodes used here, but relatively more so with SSU and LSU. Clustering of the mock community sequences indicated that the relatively more species from different fungal orders are inadvertently lumped together based on sequences of conservative domains of SSU and LSU (Suppl. material 2). Although OTU richness estimates of different barcode-primer pair combinations exhibited strong or moderate linear correlations, we caution against using SSU and LSU for metabarcoding. These genes may nonetheless be suitable for class or phylum-level analyses and inference of community phylogenies when functional assignment of taxa is not of concern.

In spite of the low taxonomic resolution of SSU and LSU, these barcodes were relatively more efficient in recovering trends in community composition in terms of greater proportion of variance explained. There are two alternative and perhaps additive explanations to this observation. First, phylogenetic niche conservatism among fungi may reinforce this pattern (Cavender-Bares et al. 2009) as taxonomic richness of fungal classes is driven by different environmental variables such as soil pH and climate (Tedersoo et al. 2014). Second, lumping of closely related rare species increases the connectance of community matrices, i.e. results in greater matrix fill. Lower connectance associated with relatively more taxon-rich data sets adversely affects the performance of many community association metrics (Olesen and Jordano 2002). We demonstrated that greater connectance may improve statistical properties of multivariate analyses by reduction of noise. Importantly, qualitative aspects of these analyses were not altered as no other variable beside vegetation became significant with removal of infrequent OTUs. These results suggest that

possibilities for optimal performance of multivariate analyses of large and sparse data sets warrant further exploration. It is important to assess the effects of connectance and removal of rare OTUs separately, because these two effects were not disentangled in our study.

Although all barcode-primer pair combinations revealed that floristic variables account for the strongest effects in fungal community composition, there were statistically significant primer biases that were not reported in previous studies (Arfi et al. 2012; Daghino et al. 2012; Blaaid et al. 2013). At the level of OTUs, the choice of primer pairs for the ITS1 barcode explained 2% of the community composition, which is less than any environmental predictor group. Tedersoo et al. (2010) found that selection of either ITS5 or ITS1F forward primers explains 15% of the community variation in ectomycorrhizal roots. Differences in performance can be explained by primer-template mismatches of the ITS1Fngs, ITS5, and ITS1ngs forward primers (Appendix 1) and by the common occurrence of an intron between ITS1Fngs/ITS5 and ITS1ngs primer sites in many ascomycete groups (Bhattacharya et al. 2001). Lower taxonomic resolution or mismatches to primers in the 3' end of the original ITS2 primer compared with optimized gITS7 and ITS3tagmix primers may have contributed to lower taxonomic richness of the ITS1 compared with ITS2 barcode. Taxa falling into Basidiomycota (except Agaricomycetes), Glomeromycota, Rozellomycota, Dothideomycetes, and Geoglossomycetes were disproportionately more OTU-rich in the ITS2 data set (Tedersoo et al. 2015b).

Metagenomics approach

Currently, the Illumina HiSeq technology enables generation of 4×10^8 DNA sequences per run. In our soil samples, fungal nuclear rDNA represented ca. 0.005% of all DNA molecules, resulting in an average sequencing depth of approx. 1000 sequences per sample ($n=16$). In spite of relatively even representation of all metagenomics sequences across samples, the coefficient of variation for the number of fungal sequences was roughly three times higher for metagenome samples compared with amplicons. Such uneven sequencing depth further complicates downstream statistical analyses.

PCR and primer biases in the amplicon data sets are well addressed, but there are also certain biases inherent to metagenomics approaches that are related to base composition and replication of DNA fragments (Gomez-Alvarez et al. 2009). Because the metagenomics sequences exhibit very low overlap across the rDNA, it is impossible to assign these sequences to OTUs and recover taxonomic richness (cf. Bengtsson et al. 2012).

Except for some minor but statistically significant taxonomic biases, the metagenomics data set covering SSU, ITS, and LSU provided highly comparable results to that of all barcodes taken together but especially to the results from ITS1 and ITS2. The metagenomics analyses confirmed that the Agaricomycetes, Eurotiomycetes, and Sordariomycetes are the key players in tropical soils of PNG. Furthermore, the ITS data set of both amplicons and metagenome exhibited a similar proportion of mycorrhizal

fungi (Tedersoo et al. 2015b). We caution that some care should be taken in interpreting our comparison of metagenomes and amplicons, because both analyses were performed in a single run, such that there may be differences between the two Illumina models, MiSeq 2×300 and HiSeq 2×250, and their chemistry.

Given the low and uneven recovery of fungal rDNA sequences and difficulties in correct taxonomic assignment (see below), metagenomics with the sole purpose of metabarcoding is clearly a waste of financial and computational resources. Enrichment of targeted molecules such as mitochondrial DNA may improve the problems associated with insufficient sequencing depth (Zhou et al. 2013). Nonetheless, vast amounts of functional data produced in the metagenomics analyses promise a new paradigm in community ecology (Zhou et al. 2015).

Analytical biases

Our *in silico* analysis of primer mismatches extends the results of Bellemain et al. (2010) and Op De Beeck (2014) by covering many additional primers and pointing to particular mismatching positions and taxa. This has enabled us to fine-tune primers for targeting the ITS region in fungi and other soil eukaryotes for cross-kingdom studies (Tedersoo et al. 2015a). We also demonstrate that there is no universally good solution for covering all fungi, because groups with rapidly evolving nuclear rDNA such as Microsporidia, certain groups within Saccharomycetes, Tulasnellaceae, and *Cantharellus* commonly escape detection by these primers for all of SSU, ITS, and LSU. Similarly, excessively long ITS barcodes and introns within barcodes may hamper amplification and sequencing of additional groups (Lindahl et al. 2013; Tedersoo et al. 2015b). Besides primer bias, unequal representation of amplicons may result from extreme GC to AT ratio (Aird et al. 2011). Although there was no evidence for length bias in our study, Ihrmark et al. (2012) demonstrated that barcode length has a strong inverse correlation with recovered abundance in a mock community based on 454 pyrosequencing technology. In particular, Archaeorhizomycetes stand out as a group with exceptionally short ITS1 and ITS2 barcodes. Because of the very strong size selection in certain cloning methods and PCR (Kanagawa 2003; Carninci et al. 2011), the unprecedented dominance of Archaeorhizomycetes in molecular studies based on cloning of long DNA fragments (Schadt et al. 2003; Castro et al. 2010) may represent an extreme example of a cloning bias.

The technical biases of sample preparation steps are poorly understood and these may be largely specific to platforms, models, and chemistry. Nonetheless, depending on the base composition and size of DNA fragments, unequal competition for adaptors may occur (Aird et al. 2011). Besides potential competitive disadvantage in ligation, the relatively longer reads in the 300–600 bp frame are less likely to be recovered because of reduced read quality in the last quarter of the sequence and problems with pairing low-quality sequences.

In addition to PCR and primer biases inherent to amplicon-based analyses, a bias related to incompleteness of reference database and uncertainty of taxonomic assignment is common to both amplicon and metagenomics data sets. A part of this so-called “identification bias” results from differential quality and abundance of reference data that may affect the probability of identification of particular taxa (Kõljalg et al. 2013). For example, reference sequences covering the full LSU are available for most early diverging lineages and the most common groups of Ascomycota and Basidiomycota, but not for “minor” groups such as Orbiliomycetes, Geoglossomycetes, and Archaeorhizomycetes. Cryptomycota and Microsporidia were poorly represented in the ITS and LSU reference data sets, because taxonomic studies in these groups are largely focused on the SSU. Similarly, many other early diverging lineages such as Zoopagomycota, Neocallimastigomycota, and Blastocladiomycota were underrepresented in the ITS reference data sets. Although the early diverging lineages represented <0.5% relative sequence abundance in all data sets, many of the OTUs not assigned to phyla probably represent one of these lineages. Another aspect of the identification bias was evident when comparing taxonomic assignments of reference databases and results from BLASTn vs. Naïve Bayesian Classifier. Combining the databases and methods enabled us to identify a vast majority of OTUs and metagenome sequences to class level. The identification bias is very difficult to quantify or explicitly test among different barcode data sets and among studies, because of confounding primer effects, PCR conditions as well as subjectively established similarity thresholds and manual re-evaluation based on previous experience and knowledge about sequence variability.

Differential representation of taxa in the reference data sets certainly plays a much greater role in metagenomics data sets targeting all genes (Korsakovskiy Pond et al. 2009) or only ribosomal DNA, because the rDNA sequences used for identification may fall into any part of the SSU, ITS, or LSU, or into intergenic spacers (IGS) that are usually not considered. Since the SSU, ITS, and LSU all include variable as well as highly conserved regions, it is very difficult to provide correct taxonomic assignments to short fragments (100-450 bases) at lower taxonomic levels. For several fungal groups, full-length sequences of LSU are lacking, which results in partial ability to identify these. Typically, fungal orders can be distinguished based on specific similarity thresholds and NBC classifications for ITS and variable domains of SSU and LSU. However, orders, classes, and even phyla can be lumped together at the same levels of similarity if the metagenomic sequence happens to cover a highly conserved region. Furthermore, our metagenomics rDNA data comprised a remarkable proportion of SSU and LSU sequences that clearly belonged to the ophisthokonts, but could not be reliably linked to animal and fungal kingdoms or minor lineages related to these groups. Therefore, the uncertainty and potential mistakes in taxonomic assignment are much more relevant to the metagenomics approach compared to DNA barcodes. To ameliorate identification biases in metagenomics data, reference databases should be updated with full-length rDNA sequences originating from genomics initiatives such as the 1000 Fungal Genomes Project (<http://1000.fungalgenomes.org>) and covering all classes of fungi and closely related organisms.

Conclusions

This study demonstrates that PCR-free metagenomics and amplicon-based approaches perform in a comparable fashion in recovering major fungal classes in spite of certain statistical differences. Within the amplicon data set, barcode-primer pair combinations differed strongly in recovering relative abundance of fungal classes and OTU richness (see also Tedersoo et al. 2015b). Nonetheless, these were all in agreement about trends in OTU richness and disentangling the key drivers of community composition. We found no evidence for reduced statistical performance in barcodes with relatively conserved sequences, but the use of conserved barcodes seriously hampers biological relevance of the data due to the inability to approximate species level (or explicitly any taxonomic level) and to assign functional categories such as trophic status (except arbuscular mycorrhizal mutualism). Considering the taxonomic resolution and primer bias, we recommend targeting the ITS2 barcode when using the current HTS technologies that permit <700 bases of high-quality reads. Due to its high taxonomic coverage, the ITS3mix series constitutes a good candidate for a forward primer and there is no better alternative than the ITS4 primer family in the beginning of LSU as a reverse primer. Both can be easily degenerated or multiplexed to address additional fungal or eukaryotic groups (Appendix 1). In strong disagreement with Waud et al. (2014), however, the ITS2 amplicon is unsuitable for studies of orchid mycorrhizal fungi due to multiple mismatches to Tulasnellaceae in primers located in the 5.8S rDNA. With technological advances in read length and quality, we advocate the use of full ITS region due to a better choice of forward primers and the presence of critical differences in either ITS1 or ITS2 in many pairs of closely related species (Blaalid et al. 2013; Kóljalg et al. 2013). Our *in silico* and *in vivo* results indicate that combining amplicons of ITS1Fngs and ITS1ngs forward primers effectively accounts for biases arising from primer mismatches and the occurrence of introns (Oja et al. 2015). In addition, the D1 and potentially D2 regions of LSU perform well in higher-level taxonomic assignment of fungi (see also Begerow et al. 2010). Reverse primers LF402 (including potential mixes) and TW13 located in LSU have strong affinities for fungi and for all eukaryotes, respectively. The use of these primers could form a strong basis for phylogenetic placement of isolates recovered from complex samples (Taylor et al. 2014).

Acknowledgements

We acknowledge C.W. Schadt and A. Rosling for raising communication about primer biases regarding Archaeorhizomycetes in HTS data sets. We thank A. Rosling for providing cultures of *Archaeorhizomyces borealis* and *A. finlayi*. We are grateful to T.M. Porter and H.T. Lumbsch for constructive editorial comments. This study received funding from Estonian Science Foundation (grants 9286, PUT0171, IUT20-30), EMP265, and FIBIR. We thank P.A. Kivistik for Illumina MiSeq and HiSeq analyses in the Estonian Genome Center.

References

- Abarenkov K, Nilsson RH, Larsson K-H, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjølner R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Ursing B, Vrålstad T, Liimatainen K, Peintner U, Kõljalg U (2010) The UNITE database for molecular identification of fungi – recent updates and future perspectives. *New Phytologist* 186: 281–285. doi: 10.1111/j.1469-8137.2009.03160.x
- Aird D, Ross MG, Chen W-S, Danielsson M, Dennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology* 12: R18. doi: 10.1186/gb-2011-12-2-r18
- Arfi Y, Buee M, Marchand C, Lévassieur A, Record E (2012) 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. doi: 10.1111/j.1574-6941.2011.01236.x
- Begerow D, Nilsson RH, Unterseher M, Maier W (2010) Current state and perspectives of fungal DNA barcoding and rapid identification procedures. *Applied Microbiology and Biotechnology* 87: 99–108. doi: 10.1007/s00253-010-2585-4
- Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kausserud H (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. *BMC Microbiology* 10: 189. doi: 10.1186/1471-2180-10-189
- Bengtsson J, Hartmann M, Unterseher M, Vaishampayan P, Abarenkov K, Durso L, Bik EM, Garey JR, Eriksson KM, Nilsson RH (2012) Megraft: a software package to graft ribosomal small subunit (16S/18S) fragments onto full-length sequences for accurate species richness and sequencing depth analysis in pyrosequencing-length metagenomes and similar environmental datasets. *Research in Microbiology* 163: 407–412. doi: 10.1016/j.resmic.2012.07.001
- Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, de Wit P, Sanchez-Garcia 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. *Methods in Ecology and Evolution* 4: 914–919. doi: 10.1111/2041-210X.12073
- Bhattacharya D, Lutzoni F, Reeb V, Simon D, Nason J, Fernandez F (2000) Widespread occurrence of spliceosomal introns in the rDNA genes of Ascomycetes. *Molecular Biology and Evolution* 17: 1971–1984. doi: 10.1093/oxfordjournals.molbev.a026298
- Blaalid R, Kumar S, Nilsson RH, Abarenkov K, Kirk PM, Kausserud H (2013) ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources* 13: 218–224. doi: 10.1111/1755-0998.12065
- Blackwell M (2011) The Fungi: 1, 2, 3 ... 5.1 million species? *American Journal of Botany* 98: 426–428. doi: 10.3732/ajb.1000298
- Borneman J, Hartin RJ (2000) PCR primers that amplify fungal rRNA genes from environmental samples. *Applied and Environmental Microbiology* 66: 4356–4360. doi: 10.1128/AEM.66.10.4356-4360.2000

- Brown SP, Veach AM, Rigdon-Huss AR, Grond K, Lickteig SK, Lothamer K, Oliver AK, Jump-ponen A (2015) Scraping the bottom of the barrel: are rare high-throughput sequences artifacts? *Fungal Ecology* 13: 221–225. doi: 10.1016/j.funeco.2014.08.006
- Carninci P, Shibata Y, Hayatsu N, Itoh M, Shiraki T, Hirozane T, Watahiki A, Shibata K, Konno H, Muramatsu M, Hayashizaki Y (2011) Balanced-size and long-size cloning of full-length, cap-trapped cDNAs into vectors of the novel lambda-FLC family allows enhanced gene discovery rate and functional analysis. *Genomics* 77: 79–90. doi: 10.1006/geno.2001.6601
- Castro HF, Classen AT, Austin ET, Norby RJ, Schadt CW (2010) Soil microbial community responses to multiple experimental climate change drivers. *Applied and Environmental Microbiology* 76: 999–1007. doi: 10.1128/AEM.02874-09
- Cavender-Bares J, Hozak KH, Fine PVA, Kembel SW (2009) The merging of community ecology and phylogenetic biology. *Ecology Letters* 12: 693–715. doi: 10.1111/j.1461-0248.2009.01314.x
- Colwell RK (2013) EstimateS: Statistical estimation of species richness and shared species from samples. Version 9. User's Guide and application published at: <http://purl.oclc.org/estimates>.
- Daghino S, Murat C, Sizzano E, Girlanda M, Perotto S (2012) Fungal diversity is not determined by mineral and chemical differences in serpentine substrates. *PLoS ONE* 7: e44233. doi: 10.1371/journal.pone.0044233
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27: 2194–2200. doi: 10.1093/bioinformatics/btr381
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 23: 3150–3152. doi: 10.1093/bioinformatics/bts565
- Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizas and rusts. *Molecular Ecology* 2: 113–118. doi: 10.1111/j.1365-294X.1993.tb00005.x
- Hildebrand F, Tadeo R, Voigt AY, Bork P, Raes J (2014) LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome* 2: 30. doi: 10.1186/2049-2618-2-30
- Hopple JS, Vilgalys R (1994) Phylogenetic relationships among coprinoid taxa and allies based on data from restriction site mapping of nuclear rDNA. *Mycologia* 86: 96–107. doi: 10.2307/3760723
- Ihrmark K, Bödeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenk J, Strid Y, Stenlid J, Brandström-Durling M, Clemmensen KE, Lindahl BD (2012) New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* 83: 666–677. doi: 10.1111/j.1574-6941.2012.01437.x
- Ishii N, Ishida S, Kagami M (2015) PCR primers for assessing community structure of aquatic fungi including Chytridiomycota and Cryptomycota. *Fungal Ecology* 13: 33–43. doi: 10.1016/j.funeco.2014.08.004
- Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Biosciences and Bioengineering* 96: 317–323. doi: 10.1016/S1389-1723(03)90130-7
- Kohout P, Sudova R, Janouskova M, Ctvrtlikova M, Hejda M, Pankova H, Slavikova R, Stajerova K, Vosatka M, Sykorova Z (2014) Comparison of commonly used primer sets for

- evaluating arbuscular mycorrhizal fungal communities: Is there a universal solution? *Soil Biology and Biochemistry* 68: 482–493. doi: 10.1016/j.soilbio.2013.08.027
- Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M (2013) Towards a unified paradigm for sequence-based identification of Fungi. *Molecular Ecology* 22: 5271–5277. doi: 10.1111/mec.12481
- Korsakovskiy P, Wadhawan S, Chairmonte E, Ananda G, Chung W-Y, Taylor J, Nekrutenko A, The Galaxy Team (2009) Windshield splatter analysis with the Galaxy metagenomic pipeline. *Genome Research* 19: 2144–2153. doi: 10.1101/gr.094508.109
- Kopylova E, Noé L, Touzet H (2012) SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28: 3211–3217. doi: 10.1093/bioinformatics/bts611
- Livermore JA, Mattes TE (2013) Phylogenetic detection of novel Cryptomycota in an Iowa (United States) aquifer and from previously collected marine and freshwater targeted high-throughput sequencing sets. *Environmental Microbiology* 15: 2333–2341. doi: 10.1111/1462-2920.12106
- Logares R, Sunagawa S, Salazar G, Cornejo-Castillo FM, Ferrera I, Sarmiento H, Hingkamp P, Ogata H, de Vargas C, Lima-Mendez G, Raes J, Poulain J, Jaillon O, Wincker P, Kandels-Lewis S, Karsenti E, Bork P, Acinas SG (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental Microbiology* 16: 2659–2671. doi: 10.1111/1462-2920.12250
- Martin KJ, Rygielwicz PT (2005) Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. *BMC Microbiology* 5: 28. doi: 10.1186/1471-2180-5-28
- Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD (2012) PANDAseq: PAired-eND Assembler for Illumina sequences. *BMC Bioinformatics* 13: 31. doi: 10.1186/1471-2105-13-31
- 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. doi: 10.1111/j.1574-6968.2009.01618.x
- Nilsson RH, Tedersoo L, Ryberg M, Kristiansson E, Hartmann M, Unterseher M, Porter TM, Bengtsson-Palme J, Walker DM, de Sousa F, Gamper HA, Larsson E, Larsson K-H, Kõljalg U, Edgar RC, Abarenkov K (in press) A comprehensive, automatically updated fungal ITS sequence dataset for reference-based chimera control in environmental sequencing efforts. *Microbes and Environments*. doi: 10.1264/jsme2.ME14121
- Oja J, Bahram M, Tedersoo L, Kull T, Kõljalg U (2015) Temporal patterns of orchid mycorrhizal fungi in meadows and forests as revealed by 454 pyrosequencing. *New Phytologist* 205: 1608–1618. doi: 10.1111/nph.13223
- Olesen JM, Jordano P (2002) Geographic patterns in plant-pollinator mutualistic networks. *Ecology* 83: 2416–2424. doi: 10.2307/3071803
- 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. *PLoS ONE* 9: e97629. doi: 10.1371/journal.pone.0097629

- Pölme S, Bahram M, Kõljalg U, Tedersoo L (2014) Global biogeography of *Alnus*-associated Frankia Actinobacteria. *New Phytologist* 204: 979–988. doi: 10.1111/nph.12962
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41: D590–D596. doi: 10.1093/nar/gks1219
- Schadt CW, Martin AP, Lipson DA, Schmidt SK (2003) Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science* 301: 1359–1361. doi: 10.1126/science.1086940
- Schadt CW, Rosling A (in press) Global diversity and geography of soil fungi: minus one widespread group. *Science*.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JS, Strez B, Thallinger GG, van Horn DJ, Weber CF (2009) Introducing Mothur: open-source, platform-independent community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75: 7537–7541. doi: 10.1128/AEM.01541-09
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An KD, Bai FY, Barreto RW, Begerow D, Bergeron MJ, 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 ZW, Griffith GW, Griffiths K, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Guo LD, Hagen F, Hambleton S, Hamelin RC, Hansen K, Harrold P, Heller G, Herrera G, Hirayama K, Hirooka Y, Ho HM, Hoffmann K, Hofstetter V, Hognabba F, Hollingsworth PM, Hong SB, Hosaka K, Houbraken J, Hughes K, Huhtinen S, Hyde KD, James T, Johnson EM, Johnson JE, Johnston PR, Jones EB, 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 SS, Martin MP, May TW, McTaggart AR, Methven AS, Meyer W, Moncalvo JM, 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 T, Ruibal C, Sarmiento-Ramirez JM, Schmitt I, Schussler A, Shearer C, Sotome K, Stefani FO, Stenroos S, Stielow B, Stockinger H, Suetrong S, Suh SO, Sung GH, Suzuki M, Tanaka K, Tedersoo L, Telleria MT, Tretter E, Untereiner WA, Urbina H, Vagvolgyi C, Vialle A, Vu TD, Walther G, Wang QM, Wang Y, Weir BS, Weiss M, White MM, Xu J, Yahr R, Yang ZL, Yurkov A, Zamora JC, Zhang N, Zhuang WY, 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 USA* 109: 6241–6246. doi: 10.1073/pnas.1117018109
- Taylor DL, Hollingsworth TN, McFarland JW, Lennon NJ, Nussbaum C, Ruess RW (2014) A first comprehensive census of fungi in soil reveals both hyperdiversity and fine-scale niche partitioning. *Ecological Monographs* 84: 3–20. doi: 10.1890/12-1693.1

- Taylor DL, McCormick MK (2008) Internal transcribed spacer primers and sequences for improved characterization of basidiomycetous orchid mycorrhizas. *New Phytol.* 177: 1020–1033. doi: 10.1111/j.1469-8137.2007.02320.x
- Tedersoo L, Bahram M, Cajthaml T, Põlme S, Hiiesalu I, Anslan S, Harend H, Buegger F, Pritsch K, Koricheva J, Abarenkov K (2015a) Tree diversity and species identity effects on soil fungi, protists and animals are context-dependent. *The ISME Journal*. [in press]
- Tedersoo L, Bahram M, Põlme S, Anslan S, Riit T, Kõljalg U, Nilsson RH, Hildebrand F, Abarenkov K (2015b) Analytical biases in microbial diversity studies. *Science*. [in press]
- Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Villarreal-Ruiz L, Vasco-Palacios A, Quang Thu P, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Ratskowsky D, Pritsch K, Riit T, Põldmaa K, Piepenbring M, Phosri C, Peterson M, Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo L, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkhan R, Dearnaley J, De Kesel A, Dang T, Chen X, Buegger F, Brearley FQ, Bonito G, Anslan S, Abell S, Abarenkov K (2014) Global diversity and geography of soil fungi. *Science* 346: 1078. doi: 10.1126/science.1256688
- Tedersoo L, Jairus T, Horton BM, Abarenkov K, Suvi T, Saar I, Kõljalg U (2008) Strong host preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA barcoding and taxon-specific primers. *New Phytologist* 180: 479–490. doi: 10.1111/j.1469-8137.2008.02561.x
- Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G, Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist* 188: 291–301. doi: 10.1111/j.1469-8137.2010.03373.x
- Tedersoo L, Smith ME (2013) Lineages of ectomycorrhizal fungi revisited: foraging strategies and novel lineages revealed by sequences from belowground. *Fungal Biology Reviews* 27: 83–99. doi: 10.1016/j.fbr.2013.09.001
- Toju H, Tanabe AS, Yamamoto S, Sato H (2012) High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PLoS ONE* 7: e40863. doi: 10.1371/journal.pone.0040863
- Turenne CY, Sanche SS, Hoban DJ, Karlowsky JA, Kabani AM (1999) Rapid identification of fungi by using the ITS2 genetic region and an automated fluorescent capillary electrophoresis system. *Journal of Clinical Microbiology* 37: 1846–1851.
- Turner S, Pryer KM, Miao VPW, Palmer JD (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *Journal of Eukaryote Microbiology* 46: 327–338. doi: 10.1111/j.1550-7408.1999.tb04612.x
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73: 5261–5267. doi: 10.1128/AEM.00062-07
- Wang X-C, Liu C, Huang L, Bengtsson-Palme J, Chen H, Zhang J-H, Cai D, Li J-Q (2015) ITS1: a DNA barcode better than ITS2 in eukaryotes? *Molecular Ecology Resources* 15: 573–586. doi: 10.1111/1755-0998.12325

- Waud M, Busschaert P, Ruyters S, Jacquemyn H, Lievens B (2014) Impact of primer choice on characterization of orchid mycorrhizal communities using 454 pyrosequencing. *Molecular Ecology Resources* 14: 679–699. doi: 10.1111/1755-0998.12229
- White TJ, Bruns TD, Lee SB, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (Eds) *PCR Protocols: A Guide to Methods and Applications*. Academic Press, New York, 315–322. doi: 10.1016/B978-0-12-372180-8.50042-1
- Zhou J, He Z, Deng Y, Tringe SG, Alvarez-Cohen L (2015) High-throughput metagenomic technologies for complex microbial community analysis: open and closed formats. *mBio* 6: e02288–14. doi: 10.1128/mBio.02288-14
- Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* 2: 4. doi: 10.1186/2047-217X-2-4

Appendix I

Sequences of widely used ITS primers, their mismatches to fungal taxa, and recommendations for metabarcoding studies.

Primers and taxa	Sequences and mismatches
I. ITS primers	
ITS1F (original: Gardes and Bruns 1993)	CTTGGTCATTAGAGGAAGTAA
ITSOF-T (Taylor and McCormick 2008)	AACTTGGTCATTAGAGGAAGT
ITSOF (Tedersoo et al. 2008)	ACTTGGTCATTAGAGGAAGT
ITS1Fngs (this study)	GGTCATTTAGAGGAAGTAA
Tulasnellaceae p.parte	***C**C*****C
<i>Cantharellus</i> p.parte	***G***T*****
Eurotiomycetes, Dothideomycetes, Sordariomycetes p.parte	***C*****
Trichocomaceae p.parte	***CC*****
Dothideomycetes p.parte	*****T****
Dothioraceae	***CC*****
Botryosphaeriaceae	***AC*****
Psoraceae	*****C*****
Valsaceae, Entomophthoromycotina	*****A*****
Chytridiomycota, Magnaporthales, <i>Suillus</i> , Paraglomerales, Mucorales, Rhiziniaceae p.parte	*****G*****
Chytridiomycota p.parte, Saccharomycetes	*****C***
Mucorales	***A**CT*****
Viridiplantae	*****TA*****G*G
Metazoa	***K*A*****W*

Primers and taxa	Sequences and mismatches
ITS5 (original: White et al. 1990)	GGAAGTAAAAGTCGTAACAAGG
Tulasnella p.parte	*****WC*****
Tremellomycetes, Ustilaginomycetes p.parte	*****M****A*****
Dothideaceae	***T*****
Saccharomycetales, Chytridiomycota p.parte	***C*****
Microsporidia, Entomophthoromycotina, Chytridiomycota p.parte	*****C*****
Metazoa	*****W*****
Viridiplantae	****G*G*****
ITS1 (original: White et al. 1990)	TCCGTAGGTGAACCTGCGG
ITS1ngs (this study)	TCCGTAGGTGAACCTGC
Sordariomycetes	****T*****A****
Microsporidia	G*W****W*****M*
Viridiplantae, Metazoa	*****
gITS7 (Ihrmark et al. 2012, used here)	GTGARTCATCGARTCTTTG
fITS7 (Ihrmark et al. 2012)	GTGARTCATCGAATCTTTG
ITS86F (Turenne et al. 1999)	GTGAATCATCGAATCTTTGAA
<i>Cantharellus</i>	*C*****G*****
Tulasnellaceae, Microsporidia	no match
<i>Antherospora</i>	*****T****T*****
Pucciniales p.parte	*****TC**GT*****
<i>Ophiostoma</i>	*C**RY*****
Eurotiales p.parte	*****G*****
<i>Cordyceps</i>	**A**CT*****A**
<i>Dipodascus</i>	*****_*****
<i>Leptosphaeria</i>	*****T*****
<i>Pichia</i>	*****G*TC*****
<i>Candida</i>	*****T**-**T*****
Neocallismatigales	*****A*****C**
Lobulomycetales	*****TA*****
<i>Acaulospora</i> , <i>Glomus intraradices</i>	*****A*****
<i>Gigaspora</i>	*****A***T*****
<i>Cokeromyces</i> , <i>Rhizopus</i>	*****G*****
Entomophthorales	*****G*T*****
Angiospermae	*Y**Y*****G*****
Gymnospermae	*****G*T*****
Metazoa p.parte	A*T*A***CA*****
Metazoa p.parte	***A*TGCA*G*CACA*K**
Straminipila	***R****R*RWY*****

Primers and taxa	Sequences and mismatches
ITS3 (original: White et al. 1990; ITS2 primer is reverse complement)	GCATCGATGAAGAACGCAGC
58A1F (Martin and Rygiewicz 2005)	GCATCGATGAAGAACGC
58A2F (Martin and Rygiewicz 2005)	ATCGATGAAGAACGCAG
ITS3-Kyo1 (Toju et al. 2012)	AHCGATGAAGAACRYAG
ITS3-Kyo3 (Toju et al. 2012)	GATGAAGAACGYAGYRAA
ITS3mix (Tedersoo et al. 2014; mix1-5 used here; consensus)	CANCGATGAAGAACGYRG
<i>Cantharellus</i>	*****Y*W*
Tulasnellaceae	no match
Mycenaceae	CY*****
<i>Amanita</i> p.parte, <i>Antrodia</i>	*****R*****
Gomphales	*****R***C****
Sebacinales p.parte	***C*****
Tremellales, <i>Yarrowia</i>	A*****
Pucciniomycetes p.parte	AY*****A**T***
<i>Puccinia</i>	R**C*****R*****
Tilletiales	C*****
Ophiocordyceps	****TA**A*****
Onygenales	A*****
Dothideomycetes p.parte	***A*****
Sordariomycetes p.parte	***A*****
Sordariales p.parte	*****G****
Pezizaceae p.parte	*****T***
Saccharomycetales p.parte	*****G*****
<i>Candida</i> p.parte	*T*****T*****
Thecaphora, Thysanophora	*T*****
Glomeraceae p.parte	*****T*****
Glomeraceae p.parte	*****A*****
Mucorales p.parte	*T*****
Mucorales p.parte, Waitea, Microbotryum	*****T*****
Chytridiomycota p.parte	***A*****
Neocallimastigomycota	*A*****
Microsporidia	*TGA*****WY*TT**
Nucleariida, Aphelidea	***A*****
Viridiplantae	*****Y*Y
Amoebozoa p.parte	*TRA*****
Amoebozoa p.parte	AT*A*****G*AT***
Amoebozoa p.parte	AA*****C**T***
Amoebozoa p.parte	**T***Y****G****T
Apusozoa p.parte	*Y*A*****
Apusozoa p.parte	A**A*****C**TG**
Alveolata p.parte	A**A*****GR*****

Primers and taxa	Sequences and mismatches
Alveolata p.parte	***C*TN*****GG*****
Alveolata p.parte	AT*A*****T***
Bacillariophyta	A**A*****
Bryophyta	***A*****
Chlorophyta p.parte	***A*****
Chlorophyta p.parte	*YG*****G****T***
Chlorophyta p.parte	*Y*N*****
Euglenozoa p.parte	ATT**T*****Y*
Euglenozoa p.parte	CAG*****G***G**
Eustigmatophyta	A**A*****
Haptophyta	***A*****
Marchantiophyta	***A*****
Platyhelminthes	*YG*****GW****
Nematoda p.parte	AGN*****
Nematoda p.parte	GGG*****R*****T
Nematoda p.parte	RGN*****R***GG**
Nematoda p.parte	*T*****A*****
Nematoda p.parte	TGG*****A***GT**
Insecta p.parte	GGG*****T
Insecta p.parte	*G*****G*****
Insecta p.parte	GGR*****C*****
Placozoa	*T*****T***
Annelida	*TG*****G*****
Mollusca	GGG*****G*****
Cnidaria	*Y*****Y***
Porifera	*T*C*****G*****
Acari p.parte	TGG*****Y**T
Acari p.parte	AA*****A***AT**T
Acari p.parte	AA*G*****A***RT**T
Acari p.parte	CA*****T***
Ichtyes p.parte	CGC*****
Oomycota p.parte	A*Y*****M***T**
Oomycota p.parte	A*****T**
Parabasalia	*****AC*T
Phacophyta	A*WA*****
Rapidophyta	A*W*****
Rhizaria p.parte	R*A*****
Rhizaria p.parte	***A*****GT*
Rhizaria p.parte	*YR*****Y*T*
Rhodophyta	RY*W*****Y***
Synurophyta	AY*A*****

Primers and taxa	Sequences and mismatches
ITS4 (original: White et al. 1990)	TCCTCCGCTTATTGATATGC
ITS4ngs (Tedersoo et al. 2014, used here)	TCCTSCGCTTATTGATATGC
Chaetothyriales	****G*****
Archaeorhizomycetes	****GC*****
Tulasnellaceae p.parte	*****G*W*A*****
Microsporidia	***S*Y*****M*****
Viridiplantae, Amoebozoa, Rhizaria	*****
Apusozoa	*****R*****
Alveolata p.parte	***K*****M*T*****
Alveolata p.parte	**T**A****A*T*****
Alveolata p.parte	**TG*****G*T*****
Alveolata p.parte	*****A*A*****
Bacillariophyta p.parte	*****A*T*****
Chlorophyta p.parte	*****A*****
Euglenozoa	*****C*R*A*****
Eustigmatophyta	*****G*T*****
Haptophyta	*****G*****
Marchantiophyta	*****G*****
Rotifera	*****K*T*****
Platyhelminthes	*****K*****
Nematoda p.parte	*****W*N*****
Nematoda p.parte	*****C*G*A*****
Insecta p.parte	*****C*C*A*****
Ichtyes	*****G*A*****
Annelida	*****A*****
Acari	*****A*A*****
Mollusca	*****Y*****
Crustacea	*****A*****
Cnidaria p.parte	***G*T***G*A*****
Cnidaria p.parte	*****K*A*****
Cnidaria p.parte	*****C*G*****
Oomycota	*****
Parabasalia	*****A**G***
Phacophyta	*****G*T*****
Rapidophyta	*****A*T*****
Rhizaria p.parte	****G*****A*****
Rhodophyta	***W*****R*W*****
Synurophyta	*****G*T*****
Nucleariida	*****C*****

Primers and taxa	Sequences and mismatches
ITS4B (Gardes and Bruns 1993)	CAGGAGACTTGTACACGGTCCAG
Hygrophoraceae, Corticiales	**A*****
Schizophyllaceae	*****A*****
<i>Armillaria</i>	*****A
Inocybaceae, Amanitaceae	*****Y*****
Psathyrellaceae	*****R*****K**GA
Amylocorticiales	*****A*****GA
Atheliales	**A*****GA
Boletales	*****R*Y*R*****K**RK
Suillineae	****A***A*****
Russulales	*****G*****
Polyporales	**A*****G*TG*
<i>Tomentella, Thelephora</i>	**A**R***K*****R***R
<i>Pseudotomentella, Phellodon, Tomentellopsis</i>	**A*G*****
<i>Hydnellum</i>	*CA*****
Hymenochaetales	**A*****R***RA
Phallomycetidae	**RRR*****R***Y**RR
<i>Clavulina, Sistotrema</i>	****A*****A*****
Ceratobasidiaceae	**A*****T*****
Sebacinales p.parte	***GA*T**A*GTGT***GA
Sebacinales p.parte	***GA*****GA
Tremellales p.parte	**A*G****RG*****
Auriculariales	**A*G*****G*
Puccinomycetes	**AC***T*****R*
Ustilaginomycetes	*****AGG*WT****GW
Dacrymycetales, Cantharellus, Tulasnella	<40% identical
Other fungal phyla, Plantae, Metazoa	<40% identical
LB-w (Tedesoo et al. 2008)	CTTTTCATCTTTCCTCACGG
<i>Cantharellus</i>	TA*****TG***
<i>Sistotrema</i>	*****G***
<i>Tulasnella</i>	*****C*****G***
Coniophoraceae	*****C*****
Ustilaginomycetes	*****A***T*
Ascomycota	*****GA***TC
Saccharomycetes	*****W***W*
Glomeromycota, Mycoromycota, Chytridiomycota	*****
Gigasporaceae	*****A***T*
Viridiplantae	*****G***

Primers and taxa	Sequences and mismatches
II. LSU primers	
LR0R (Hopple and Vilgalys 1994)	ACCCGCTGAACTTAAGC
LR0Rngs (this study)	ACSCGCTGAACTTAAGC
Tulasnellaceae	**G*C**NR*Y*****
Chaetothyriales	**G*****
Archaeorhizomycetes	***GC*****
Microsporidia	MMMKS*Y**R*****
Viridiplantae	*****T*****
LF402Fmix1 (this study; LF402 is a reverse complement)	GTGAAATTGYTRAAAGGGAA
LF402Fmix3 (this study)	GTGAAATTGTCAAAAGGGAA
Most fungi	*****TTG*****
<i>Cantharellus</i>	*****CG**A***
Tulasnellaceae p.parte	*****Y*GTR****
Agaricaceae, Boletaceae	*****C*****
Ceratobasidiaceae	*****T****
Cystofilobasidiaceae, Corticiaceae	*****A*****
Ustilaginaceae	*****CCA*****
Schizosaccharomycetaceae	*****C***R****
Glomerellaceae, <i>Verticillium</i> , Dothideomycetes	*****A*****
<i>Candida</i> p.parte	*****A***T*CW***
<i>Falcocladium</i>	*****C*A*****
Mucoraceae p.parte	*****A*****
Neocallimastigales, Chytridiomycota	*****CR*****
Viridiplantae	T*****C*GG*****
TW13 (original: T.J. White, unpublished)	GGTCCGTGTTTCAAGACG
LR3 (original: Hopple and Vilgalys 1994; LR3R is reverse complement)	CCGTGTTTCAAGACGGG
Saccharomycetales p.parte	****A*****
Microsporidia, Viridiplantae, Metazoa	*****
Dictyostelids	RR**YR*R***T****TA**
LR5-Fung (Tedersoo et al. 2008)	CGATCGATTGCACGTCAGA
Tulasnellaceae	*****Y*****
Cordycipitaceae p.parte	*****C*****
Straminipila, Metazoa	*****
Viridiplantae, Alveolata, Rhizaria	***A*****T
LR5 (original: Hopple and Vilgalys 1994)	TCCTGAGGAACTTCG
TW14 (T.J. White et al. unpublished)	GCTATCCTGAGGAACTTC
Microsporidia, Metazoa, Viridiplantae	*****
<i>Candida</i> p.parte	*****A*****
Apicomplexa p.parte	**G*****A**TCT****
Straminipila	*****Y*****

Primers and taxa	Sequences and mismatches
III. SSU primers	
SSU515F (original: Turner et al. 1999)	<u>GTGCCAGCANCCGCGGTAA</u>
SSU515Fngs (this study)	<u>GCCAGCAGCCGCGGTAA</u>
Saccharomycetes, Pezizomycotina p.parte (I, intron site)	*T*****
Archaeorhizomycetes	*****C*****
Microsporidia	*****T*****
Viridiplantae, Metazoa	*****
Euk742R (this study)	AAATCCAAGAATTTACCTCT
Many fungal groups	NR*****
<i>Cantharellus</i>	TGG**ACT*G*****C**
Tulasnellaceae	GGR**ANM*R*****Y**
Baeomycetales p.parte	*****Y***
Eurotiales p.parte	NR*****Y*Y*
Harpellales	TGG*T***T*****A**
Dothideomycetes p.parte	CG*****G*****C**
Lecanorales	RW**NMN*****Y***
Saccharomycetales, Hypocreales	NNR**Y*****
Chytridiomycota, Neocallimastigomycota, Blastocladiomycota	****Y*****
SSU817F (original: Borneman and Hartin 2000)	TTAGCATGGAATAATRRATAGGA
<i>Cantharellus</i>	*****G***CSG*WWC***
Tulasnellaceae	*****RCWN*TKGACN
Sebacinales p.parte	*****G*****
Gomphales	*****C*****
Phallales	*****R**
Agaricostilbales p.parte	AC*****
Dacrymycetales p.parte	*****K*****
Ustilaginales p.parte	*****C*****
Pucciniales p.parte	*****C*A**
Coryneliales p.parte	*****A**
Helotiales p.parte	*****R*****N*N*M
Microascales p.parte	WY*****
Pleosporales p.parte	NY**W*****R
Saccharomycetales p.parte	NN*T*****G*CAGG*CC*YT
Saccharomycetales p.parte	NN*****C**T*--**
Saccharomycetales p.parte	WY*****N*****
Urocystidales	*****N***Y***
<i>Ramicandelaber, Falciformispora</i>	*****G*****
Zoopagomycota	*****C*****
Mucorales p.parte	*****N**
Entomophthoromycota	*****K*N*W
Chytridiomycota	*Y*****Y*****
Microsporidia, Viridiplantae	no match

Primers and taxa	Sequences and mismatches
SSU1196R (original: Borneman and Hartin 2000)	TCTGGACCTGGTGAGTTTCC
SSU1196Rngs (this study)	TCTGGACCTGGTGAGTTT
<i>Cantharellus</i>	**CT*****T*
Tulasnellaceae	**CT*T*****GT*
Tremellales p.parte, Chytridiomycota	*****A*****
Saccharomycetes p.parte	*****C*GT*
Pezizales p.parte	*****A**T*
Sordariales p.parte	**C****C*****
Glomeromycota p.parte	*****R*****T*
Microsporidia	NNNNNR**N**R**R*KT*
Viridiplantae	*****A*****
IV. Primer recommendations	
Recommended primer mixes for the ITS1F family	
ITS1Fngs-Mix1 (Fungi)	GGTCATTTAGAGGAAGTAA
ITS1Fngs-Mix2 (Tulasnellaceae)	GGCCATTTAGAGGAAGTAC
ITS1Fngs-Mix3 (Saccharomycetes)	GGTCATTTAGAGGAAGTAA
ITS1Fngs-Mix4 (various groups)	GGTCGTTTAGAGGAAGTAA
ITS1Fngs-Mix5 (Mucorales)	GGCTATTTAGAGGAAGTAA
Recommended primer mixes for the ITS1 family	
ITS1ngs-Mix1 (Most eukaryotes)	TCCGTAGGTGAACCTGC_
ITS1ngs-Mix2 (Sordariomycetes)	TCCGTTGGTGAACCAGC_
Recommended ITS1 and full ITS forward primer mixes for fungi	
ITS1Fngs (except SSU 5' intron containing groups)	GGTCATTTAGAGGAAGTAA
ITS1ngs (except Sordariomycetes)	TCCGTAGGTGAACCTGC
Recommended forward primer mixes for ITS2 barcode	
ITS3-Mix1 (Fungi)	CATCGATGAAGAACGCAG_
ITS3-Mix2 (Chytridiomycota)	CAACGATGAAGAACGCAG_
ITS3-Mix3 (Sebacinales)	CACCGATGAAGAACGCAG_
ITS3-Mix4 (Glomeromycota)	CATCGATGAAGAACGTAG_
ITS3-Mix5 (Sordariales)	CATCGATGAAGAACGTGG_
Recommended reverse primers for ITS2 and full ITS	
ITS4-Mix1 (Fungi)	TCCTCCGCTTATTGATATGC
ITS4-Mix2 (Chaetothyriales)	TCCTGCGCTTATTGATATGC
ITS4-Mix3 (Archaeorhizomycetes)	TCCTCGCCTTATTGATATGC
ITS4-Mix4 (Tulasnellaceae)	TCCTCCGCTGAWTAATATGC
ITS4-Euk (all eukaryotes)	TCCTSSGCTTANTDATATGC
Recommended LF402 mixes for fungi	
LF402f_mix1 (Fungi)	TTCCCTTTYARCAATTTAC
LF402f_mix2 (Ceratobasidiaceae)	TTCCATTTCAACAATTTAC
LF402f_mix3 (Chytridiomycota)	TTCCCTTTTGACAATTTAC
LF402f_mix4 (Tulasnellaceae)	TTCCCYACCRAAATTTAC
LF402f_mix5 (Cantharellus)	TTCTCCGTCAACAATTTAC

Supplementary material 1

Table S1. Characteristics of soil samples.

Authors: Leho Tedersoo, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, Veljo Kisand, R. Henrik Nilsson, Falk Hildebrand, Peer Bork , Kessy Abarenkov

Data type: measurement

Explanation note: Characteristics of soil samples used in this study.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 2

Table S2. Taxonomic composition and clustering of the mock community sample.

Authors: Leho Tedersoo, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, Veljo Kisand, R. Henrik Nilsson, Falk Hildebrand, Peer Bork , Kessy Abarenkov

Data type: measurement

Explanation note: Taxonomic composition and clustering of the mock community sample.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 3

Table S3. Data set of the SSU V4 and V5 barcodes.

Authors: Leho Tedersoo, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, Veljo Kisand, R. Henrik Nilsson, Falk Hildebrand, Peer Bork , Kessy Abarenkov

Data type: data set

Explanation note: Data set of the SSU V4 and V5 barcodes.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 4

Table S4. Data set of the ITS1 barcode.

Authors: Leho Tedersoo, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, Veljo Kisand, R. Henrik Nilsson, Falk Hildebrand, Peer Bork, Kessy Abarenkov

Data type: data set

Explanation note: Data set of the ITS1 barcode.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 5

Table S5. Data set of the ITS2 barcode.

Authors: Leho Tedersoo, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, Veljo Kisand, R. Henrik Nilsson, Falk Hildebrand, Peer Bork, Kessy Abarenkov

Data type: data set

Explanation note: Data set of the ITS2 barcode.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 6

Table S6. Data set of the LSU D1, D2, and D3 barcodes.

Authors: Leho Tedersoo, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, Veljo Kisand, R. Henrik Nilsson, Falk Hildebrand, Peer Bork, Kessy Abarenkov

Data type: data set

Explanation note: Data set of the LSU D1, D2, and D3 barcodes.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Supplementary material 7

Table S7. Taxonomic classification of the rDNA of fungal.

Authors: Leho Tedersoo, Sten Anslan, Mohammad Bahram, Sergei Põlme, Taavi Riit, Ingrid Liiv, Urmas Kõljalg, Veljo Kisand, R. Henrik Nilsson, Falk Hildebrand, Peer Bork , Kessy Abarenkov

Data type: taxonomic data

Explanation note: Taxonomic classification of the rDNA of fungal shotgun metagenome.

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Cloning of ribosomal ITS PCR products creates frequent, non-random chimeric sequences – a test involving heterozygotes between *Gymnopus dichrous* taxa I and II

Karen W. Hughes¹, Samuel D. Morris², Ana Reboredo-Segovia³

1 Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37920 **2** 130 Ervin Dr., Nashville, TN 37075 **3** Erasmus Mundus Master's Program in Tropical Biodiversity and Ecosystems, Université Pierre et Marie Curie - UPMC Paris, France

Corresponding author: Karen W. Hughes (khughes@utk.edu)

Academic editor: R. Henrik Nilsson | Received 16 April 2014 | Accepted 6 June 2015 | Published 19 June 2015

Citation: Hughes KW, Morris SD, Reboredo-Segovia A (2015) Cloning of ribosomal ITS PCR products creates frequent, non-random chimeric sequences – a test involving heterozygotes between *Gymnopus dichrous* taxa I and II. MycoKeys 10: 45–56. doi: 10.3897/mycokeys.10.5126

Abstract

Gymnopus dichrous exists in the southern Appalachians (USA) as two distinct entities with essentially identical nuclear ribosomal ITS1 sequences but differing ITS2 and LSU sequences (for convenience, called *G. dichrous* I and II). F₁ ITS heterozygotes between the two are routinely collected from nature. Cloning of ITS PCR products from F₁ heterozygotes produced sequences of both parental haplotypes but also numerous chimeric sequences (21.9%). The location of template switching was non-random leading to recovery of the same chimera several times and the chimeric region varied from 45bp to 300bp. By comparison, single-basidiospore isolates from heterozygote F₁ fruitbodies showed no recombinant haplotypes within the ITS + LSU span and clones derived from P₁ homozygotes were identical to the P₁ parent. Thus, chimeric sequences are likely an artifact of the PCR-cloning process and not a consequence of natural recombination events found in nature, nor are they due to hidden existing variation within the ribosomal repeat. Chimeras and PCR-induced mutations are common in cloned PCR products and may result in incorrect sequence information in public databases.

Key words

Basidiomycete fungi, chimera, hybrid, monokaryons, recombination, ribosomal repeat

Introduction

Sequence chimeras are common when pooled DNA s are co-amplified by a PCR process (Edgar et al. 2011; Judo et al. 1998; Jumpponen 2007; Meyerhans et al. 1990; Odelberg et al. 1995; Qiu et al. 2001a; Smyth et al. 2010; Tedersoo et al. 2014; Wang and Wang 1997). There has been speculation that chimeras may be the result of incomplete extension of PCR products which subsequently act as primers for the next amplification cycle and, in fact, there seems to be a reduction in chimeric PCR products when extension times are increased (Meyerhans et al. 1990; Qiu et al. 2001a; Smyth et al. 2010; Thompson et al. 2002) or by optimizing the PCR protocol (Qiu et al. 2001b; Wang and Wang 1997). Odelberg et al. (1995) however, demonstrated that chimeras can be generated in a single round of PCR amplification in the absence of heat denaturation and re-annealing which suggests that some polymerase template switching may occur. They demonstrated that template switching was reduced several fold (but not eliminated) by fixing templates to streptavidin magnetic beads.

The overall frequency of chimeras in published studies is unknown. In a study of fungal ITS amplicons derived from soil samples using a PCR/cloning process (O'Brian et al. 2005), 5% were chimeric as determined by an ascomycete/basidiomycete discontinuity between ITS1 and ITS2 regions. Smaller chimeras between related taxa would not have been detected in this study [see later detection of a chimera in this data set by Ryberg et al. (2008)]. For nSSU sequences, the proportion of chimeric sequences can be extensive (Ashelford et al. 2005; Ashelford et al. 2006; Fonseca et al. 2012; Haas et al. 2011; Wang and Wang 1997) leading to 'false diversity estimates and false novel taxa. Fonseca et al (2012) demonstrated that nuclear SSU chimeras were produced at high levels during the PCR-process when mixed templates were present. Interestingly, their results with a nematode population demonstrated that chimera formation is higher in species-diverse PCR pools than in pools that are genetically less diverse but that the breakpoints were in regions of sequence similarity.

Most programs for checking and removing chimeric sequences were designed for 16S ribosomal sequences. DECIPHER (Wright et al. 2012) (<http://decipher.cee.wisc.edu/index.html>) for bacterial 16S sequences searches for sets of short fragments that are uncommon in the phylogenetic group where the sequence is classified, but frequent in other phylogenetic groups. This depends on a robust pre-existing data set. UCHIME (Edgar et al. 2011), a chimera check program, is best when two known parental sequences and a high-quality reference database that is chimera free are available. Bellerophon uses a partial treeing analyses to detect 16S chimeras (Huber et al. 2004). Other chimera checking programs are available [see Fonseca et al. (2012) for discussion].

Nilsson et al. (2012) note the high incidence of ITS chimeras in public ITS databases and suggest a mechanism for identifying them by Blast mismatches between the ITS1 and ITS2 regions and by exploring long branches in phylogenetic trees. They comment that the most frequent site for exchange between two similar templates in a PCR reaction is in the first part of the highly conserved 5.8S segment. An open source chimera checker for the ITS region has been developed and is available at <http://www>.

emerencia.org/chimerachecker.html (Nilsson et al. 2010). Using chimera checker to evaluate 12 300 sequences, 1.5% were identified as chimeric sequences. To better facilitate ITS-based molecular identification of fungi for the scientific community, the UNITE database was established to provide reliable documented ITS sequences for the fungal community (Kõljalg et al. 2013; Nilsson et al. 2015) and a reference set of ITS sequences, each representing a species hypothesis is available in several formats including UCHIME (Nilsson et al. 2015), available at <https://unite.ut.ee/repository.php#uchime>.

With the establishment of the ribosomal ITS as the fungal barcode (Schoch et al. 2012), and the identification of highly heterozygous fungal ITS sequences that require cloning to resolve, it becomes increasingly important to understand the consequences of cloning PCR products from mixed templates including cloned PCR products generated during environmental sampling. Divergent *Gymnopus dichrous* ITS2 haplotypes provide an appropriate experimental system with which to explore this issue.

Gymnopus dichrous is a small, saprobic mushroom commonly found on oak bark and other woody debris in mid-summer in the southern Appalachian Mountains (USA). ITS sequencing identified two ITS subgroups of this mushroom called for convenience *G. dichrous* I and II. *Gymnopus dichrous* I and II differ in the ITS2 region (10% divergence) but there are no consistent or significant bp differences between *G. dichrous* I and II ITS1 or 5.8S regions (average divergence = 0.29% between the *G. dichrous* I and II ITS1 region, 0% in the 5.8s region). Homozygous collections for *G. dichrous* I and *G. dichrous* II were collected in the southern Appalachian Mountains and were designated as parental genotypes (P_1). Fruitbodies that were ITS hybrids between *G. dichrous* I and *G. dichrous* II were also been collected and were designated as F_1 hybrids (first filial generation as used in standard genetic crosses). For F_1 hybrids, several indels in the ITS2 region obscured electropherograms and prevented recovery of the parental ITS sequences during Sanger sequencing. Cloning of the ITS1-5.8S-ITS2 PCR product was required to recover individual contributing haplotypes, however, a significant portion of recovered haplotypes were chimeric sequences.

Cloned ITS sequences were compared to P_1 (parental/ homozygote) ITS sequences of *Gymnopus dichrous* I and II to identify chimeric and non-chimeric sequences. Below, we examine chimeras derived from cloned *G. dichrous* heterozygotes and show that they can be small, frequent and non-random. We also provide evidence that putative chimeras were not due to natural meiotic recombination or variation in the ribosomal repeat.

Methods and materials

Collections. *Gymnopus dichrous* and *G. subnudus* are often collected as the same entity and are morphologically difficult to separate. Both are variable in morphology. Putative *Gymnopus dichrous* collections were made in the field using known morphological and environmental characteristics [e.g., small (ca. 5 cm in height) brown mushrooms, often

with a darker, compressed, stem base and growing on wood, usually but not exclusively on oak bark]. Of 116 collections of putative *G. dichrous*, 16 were *G. dichrous* I-II hybrids (Table 1). Collections are archived in TENN-Fungi.

Single basidiospore isolation. Single-basidiospore isolates (SBIs) were obtained from fresh spore drops as described in Gordon and Petersen (1992). Monokaryon status of SBI cultures was determined microscopically by lack of clamp connections.

PCR and Cloning procedures. Cloning was carried out using Promega's pGEM-T easy kit and M109 Competent cells according to manufacturer's directions (Promega). Sanger sequencing of ITS-cloned products was performed as described in Hughes et al. (Hughes et al. 2009).

PCR of the nuclear ribosomal ITS area was performed using primers ITS1F (Gardes and Bruns 1993) and ITS4 (White et al. 1990) for all collections in this study. PCR parameters for ITS amplification were 3 min at 94 °C followed by 34 cycles of 94 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min and a final extension of 72 °C for 3 min. Each 50 µl PCR reaction contained 24.25 µl sterile ddH₂O, 10 µl of 5X PCR buffer (Promega Corporation, 2800 Woods Hollow Road, Madison, WI 53711 USA), 2.5 µl 5% DMSO (Sigma-Aldrich Company, St. Louis, MO, USA), 6 µl of 25 mM Mg = 3 mM (Promega), 4 µl of 100 mM dNTP mix (Promega), 1 µl each of two primers (10 µM) and 0.25 µl *Taq* polymerase (Promega). Sequencing of the 5' end of the nuclear ribosomal RNA large subunit gene was performed for single spore isolates only and used primers LR0R and LR5 (Vilgalys and Hester 1990). Parameters for LSU amplification are the same as for ITS except the extension time at 72 °C was 1.5 min.

Identification of template switching regions. There are 25 regions of sequence mismatch within *G. dichrous* I and *G. dichrous* II (Fig. 1, red text) which were used to identify apparent template switching. The first difference at ITS2 position 15 and the 1bp indel at position 24 were used to establish whether the 5' end of ITS2 represented *G. dichrous* I or *G. dichrous* II. Base pair 31 (A or G) is variable within the *G. dichrous* I population and when it is an adenine residue, it can be used as an additional marker. Clones exhibiting template switching were identified by an observed sequence change from *G. dichrous* I to *G. dichrous* II (or vice-versa) between bases 25 and 332. Four discontinuities were observed in the data set between bp15 and bp25. In two cases, bp25 and an adenine/guanine base pair at bp31 were used to identify the 5' end of the clone as *G. dichrous* I or II. In the other two cases, bp25 and bp71 were used to identify the 5' end of the clone. The discontinuities may be due to template switching or to PCR generated mutation. We note that rare PCR-induced base pair mutations could affect determination of the correct template switching point.

DNA folding. Potential DNA folding of the ITS2 region for *G. dichrous* I and II exemplars was estimated at 72 °C (extension phase of PCR) using the MFOLD web server (<http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form>) with a MgCl setting of 3 mM (Zuker 2003) and 0 mM Na.

Chi square analysis. The ITS 2 region was divided into 25 segments of varying length between bases 25 and 332. Each segment was flanked by a sequence difference between *G. dichrous* I and II that was informative for diagnosing template switching.

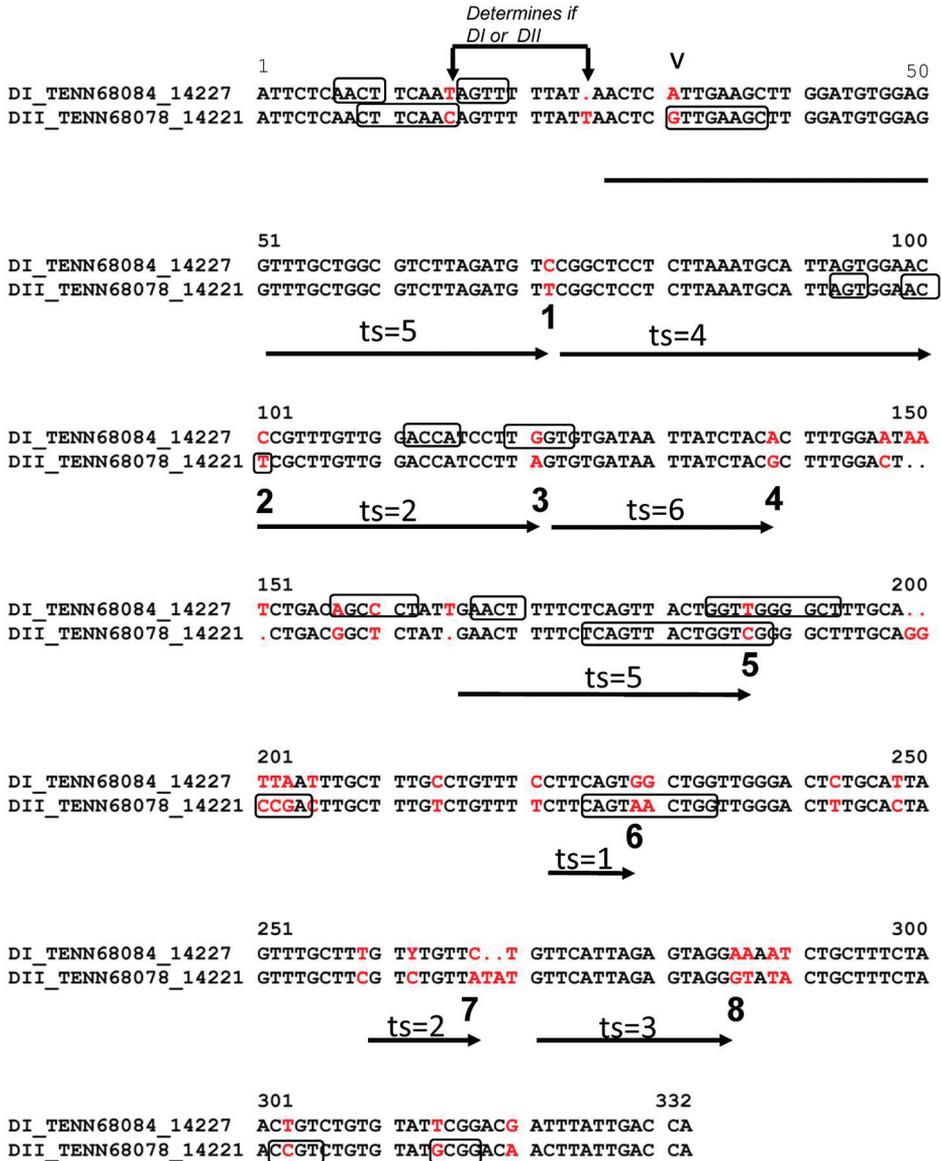


Figure 1. The ITS2 region of two haplotypes of *Gymnopus dichrous*. Haplotypes DI (represented by TENN68084) and DII (represented by TENN68078). The TC pair at position 15 and the indel at position 25 were used to determine which the haplotype was represented by the 5' end of a cloned sequence. Bases in red are points where DI and DII haplotypes differ in sequence and were used to determine if template switching had occurred in a cloned PCR product. Eight base pairs at which template switching can be detected are indicated by numbers 1-8. The possible area in which template switching (ts) could have occurred is indicated by vertical arrows and the number of observed template switching events is given above the vertical arrow. Bases that may be involved in intra-strand base pairing as determined by MFOLD are outlined with black boxes. Ambiguity codes indicate intraspecific variation.

Table 1. Distribution of ITS sequences from clones of *Gymnopus dichrous* I-II heterozygotes.

Collection number of heterozygote	Herbarium Number ¹	GenBank Numbers	Number of clones	<i>G. dichrous</i> I (# recovered)	<i>G. dichrous</i> II (# recovered)	<i>G. dichrous chimeras</i> (# recovered)
WS9223	DEWV9223		7	2	2	3
WRWV03-989	DEWV4744		6	3	3	0
WRWV04-322	DEWV5833		6	2	3	1
10454	TENN58148		8	4	2	2
10455	TENN58149		2	2	0	0
11785	TENN60014	JF313671- JF313672	4	3	1	0
11814	TENN60027	JF313673- JF313677	5	2	2	1
13361	TENN61624	JF313678- JF313693	13	5	6	2
14061	TENN67807		8	2	4	2
14094	TENN67843		8	2	3	3
14111	TENN67859		8	3	4	1
14226	TENN68083		8	2	3	3
14230	TENN68090		8	1	4	3
14237	TENN68092		7	1	4	2
14247	TENN68105		16	4	9	3
14248	TENN68106		14	6	6	2
Totals			128	44 = 34.37%	56 = 43.75%	28 = 21.88%

¹TENN-F-XXXXXX is abbreviated as TENNXXXXXX

For each segment, the number of template switching events was recorded, ranging from zero to six. These constituted observed values. Expected values were based on a null hypothesis of random template switching (each base has an equal probability of template switching).

Results and discussion

The proportion of ITS chimeras obtained from PCR amplification of the ITS regions of *G. dichrous* I-II heterozygotes is given in Table 1. Of 128 clones from F₁ heterozygotes, 28 (21.9%) were chimeric in the ITS2 region. The number of clones representing non-chimeric ITS sequences was unequally distributed between *G. dichrous* I and II. From F₁ heterozygotes, 34.4% of the clones were *G. dichrous* I sequences and 43.75% were *G. dichrous* II sequences.

The distribution of template switch points resulting in chimeras was not random along the ITS2 region (Chi square = 35.72, P<0.05). Of 25 possible diagnostic base pair sites, template switching was observed between only 8 points (Fig. 1) and the most frequent template switching (six events) occurred within a very short span of bases, bases 122-138 creating 6 identical chimeric sequences. Two regions of 5 tem-

plate switching events were also observed (Fig. 1). It should be noted that the exact point at which template switching occurred cannot be determined in regions where *G. dichrous* I and II have identical sequence but is between the last 5' diagnostic base pair difference between *G. dichrous* I and II and the diagnostic base pair at which template switching occurred (Fig. 1, horizontal arrows). The possibility of identical chimeras occurring in GenBank and thus being interpreted as valid taxa was noted by Nilsson et al. (2012).

The non-random nature of template switching would suggest that some stable mechanism influences template switching. We investigated the possibility that secondary structure formation during the PCR process might lead to non-random chimera formation, perhaps by briefly stalling *taq* polymerase transcription at the point of secondary folding and allowing template switching. Ribosomal ITS2 RNA is known to have secondary structure at normal cellular temperatures and conditions (Joseph et al. 1999; Krüger and Gargas 2004; Krüger and Gargas 2008). We wondered if the formation of DNA secondary structure in the ITS2 region might affect chimera formation. Using MFOLD (Zuker 2003) we examined secondary structure of ITS2 DNAs at 72 °C (extension phase). At 72 °C, MFOLD predicted stable secondary structure in both *G. dichrous* I and *G. dichrous* II templates. For *G. dichrous* I, 4 possible folding configurations were reported at increasing levels of free energy (dG). The folding configuration with the lowest free energy is mapped to the sequence in Fig. 1. For *G. dichrous* II, a single folding configuration was reported. This is also mapped to Fig. 1. We note that factors other than the ITS2 sequence itself influence RNA folding.

The most frequent template switching occurred in a region between bases 122 and 138. This region overlaps and follows a small area of folding (a 4bp neck and 4bp loop) in *G. dichrous* I templates. The region between 140 and 200 is involved in complex folding patterns that are not consistent from model to model but present in all models. Five template switching events occur in this region. Between bases 200 and the end of the template at base 332, there is no consistently predicted secondary structure and fewer template switching events. Thus there is a loose agreement between secondary structure and regions of template switching but we cannot conclude cause and effect.

The size of detectable chimeras varied from chimeras occurring at the 5' end of the ITS2 sequence (approximately 300bp) to those occurring near the 3' end (45bp). Chimeras occurring between bp15 and bp25 would not have been recorded as such by our procedure but may have occurred (4 discontinuities may be due to template switching or to PCR generated mutation-see methods). We note that there is an area of secondary structure which overlaps bp15 and could be involved in template switching.

Areas where *G. dichrous* DI and DII differ extensively including indels (bases 147–151, 198–205) do not seem to be involved in template switching. This has been noted in other studies as well (Fonseca et al. 2012; Haas et al. 2011).

To evaluate whether cloning simply uncovered existing variation in the ribosomal repeat region (Lindner and Banik 2011), we sequenced the ITS plus LSU region of single basidiospore isolates (SBIs) derived from F1 heterozygotes (Table 2). SBIs of F₁ spores were either *G. dichrous* I or II and did not show any evidence of meiotic

Table 2. Concatinated single basidiospore isolate (SBI) ITS + LSU sequences.

Collection number of heterozygote	Herbarium Number	Number of SBIs	<i>G. dichrous</i> I	<i>G. dichrous</i> II	Number of meiotic recombinants In ITS+LSU
F₁ <i>G. dichrous</i> I-II heterozygotes					
14111	TENN67859	11	5	6	0
14226	TENN68083	8	5	3	0
14237	TENN68092	5	1	4	0
Totals			11 (44%)	13 (52%)	0
P₁ <i>G. dichrous</i> I homozygotes					
13716	TENN65070	10	10	0	0
14232	TENN68091	10	10	0	0
P₁ <i>G. dichrous</i> II homozygotes					
Braaton	No specimen	4	0	4	0

recombination between ITS + LSU types I and II. We further examined SBI sequences from two collections of *G. dichrous* I (Table 2) and found no evidence that any *G. dichrous* II sequence elements were present in the ribosomal repeat. Finally, we cloned PCR products from three collections of *G. dichrous* I (TENN68152–8 clones, TENN67834–7 clones, TENN69091–8 clones) and again found no evidence for the presence of *G. dichrous* II sequence elements. We conclude that there is no current evidence in *G. dichrous* that cloning is recovering existing intragenomic variation in the ribosomal repeat but cannot exclude that possibility. Results reported as intragenomic variation in *Laetiporus cincinnatus* by Lindner and Banik (2011) could be due to differences in one or more ribosomal repeats but could also be explained if the sampled fruitbody was a hybrid and the two parental haplotypes and resulting chimeras were recovered.

Conclusions

Chimeras are common in cloned PCR products and tend to obscure contributing parental haplotypes, thus potentially creating errors in DNA sequence repositories. In this study, we show:

1. Template switching is non-random. Of 25 possible markers where ITS2 sequences of *G. dichrous* I and II differ, only 8 show template switching and template switching is higher in specific regions of the ITS2 sequences. The non-random nature of chimeras could lead to the misinterpretation of chimeras as parental haplotypes when the same chimera is recovered multiple times.
2. There is a loose correlation between areas predicted to form secondary structure and regions where template switching is high. We conclude that formation of secondary structure may affect template switching but speculate that secondary

structure formation could either enhance or repress template switching, depending on location and the size of the stem-loop structure.

3. Chimeras occurring near the end of a template may be short and thus not easily detected.
4. Chimeras are not due to recovery of underlying variability in the ribosomal repeat in this system. The origins of chimeras remain obscure and may be due to multiple factors.
5. Chimera control should be exercised in environmental sampling studies and taxonomic studies wherever possible in order to minimize problems with persistent errors in sequence data repositories.

Acknowledgements

This study was supported by NSF DEB1144974 to R. H. Petersen, KWH, and Brian O'Meara. We thank reviewers Henrik Nilsson, Martin Ryberg Leho Tedersoo, Shawn Brown and an anonymous reviewer for their careful reviews, comments and suggestions.

References

- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At Least 1 in 20 16S rRNA Sequence Records Currently Held in Public Repositories Is Estimated To Contain Substantial Anomalies. *Applied and Environmental Microbiology* 71: 7724–7736. doi: 10.1128/aem.71.12.7724-7736.2005
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2006) New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl Environ Microbiol*, 5734–5741. doi: 10.1128/AEM.00556-06
- Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27(16): 2194–2200. doi: 10.1093/bioinformatics/btr381
- Fonseca VG, Nichols B, Lallias D, Quince C, Carvalho GR, Power DM, Creer S (2012) Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. *Nucleic Acids Research* 40: e66. doi: 10.1093/nar/gks002
- Gardes M, Bruns TD (1993) ITS primers with enhanced specificity for basidiomycetes – application to the identification of mycorrhizae and rusts. *Mol Ecol* 2: 113–118. doi: 10.1111/j.1365-294X.1993.tb00005.x
- Gordon SA, Petersen RH (1992) Interbreeding populations of some *Marasmius* species. *Mycologia* 84: 204–208. doi: 10.2307/3760251
- Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methé B, DeSantis TZ, Consortium THM, Petrosino JF, Knight R, Birren BW (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research* 21: 494–504. doi: 10.1101/gr.112730.110

- Huber T, Faulkner G, Hugenholtz P (2004) Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* 20: 2317–2319. doi: 10.1093/bioinformatics/bth226
- Hughes KW, Petersen RH, Lickey EB (2009) Using heterozygosity to estimate a percentage DNA sequence similarity for environmental species delimitation across basidiomycete fungi. *New Phytologist* 182: 795–798. doi: 10.1111/j.1469-8137.2009.02802.x
- Joseph N, Krauskopf E, Vera MI, Michot B (1999) Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast. *Nucleic Acids Research* 27: 4533–4540. doi: 10.1093/nar/27.23.4533
- Judo MSB, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Research* 26: 1819–1825. doi: 10.1093/nar/26.7.1819
- Jumpponen A (2007) Soil Fungal Communities Underneath Willow Canopies on a Primary Successional Glacier Forefront: rDNA Sequence Results Can Be Affected by Primer Selection and Chimeric Data. *Microbial Ecology* 53: 233–246. doi: 10.1007/s00248-004-0006-x
- Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Põldmaa K, Saag L, Saar I, Schüßler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson K-H (2013) Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology* 22: 5271–5277. doi: 10.1111/mec.12481
- Krüger D, Gargas A (2004) The basidiomycete genus *Polyporus* – an emendation based on phylogeny and putative secondary structure of ribosomal RNA molecules. *Feddes Repertorium* 115: 530–546. doi: 10.1002/fedr.200311052
- Krüger D, Gargas A (2008) Secondary structure of ITS2 rRNA provides taxonomic characters for systematic studies — a case in *Lycoperdaceae* (Basidiomycota). *Mycological Research* 112: 316–330. doi: 10.1016/j.mycres.2007.10.019
- 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. doi: 10.3852/10-331
- Meyerhans A, Vartanian J-P, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Research* 18: 1687–1691. doi: 10.1093/nar/18.7.1687
- Nilsson RH, Tedersoo L, Abarenkov K, Ryberg M, Kristiansson E, Hartmann M, Schoch C, Nylande rJ, Bergsten J, Porter T, Jumpponen A, Vaishampayan P, Ovaskainen O, Hallenberg N, Bengtsson-Palme J, Eriksson K, Larsson K, Larsson E, Kõljalg U (2012) Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *Myckeys* 4: 37–63. doi: 10.3897/myckeys.4.3606
- Nilsson RH, Tedersoo L, Ryberg M, Kristiansson E, Hartmann M, Unterseher M, Porter TM, Bengtsson-Palme J, Walker DM, de Sousa F, Andres Gamper H, Larsson E, Larsson K-H, Kõljalg U, Edgar RC, Abarenkov K (2015) A Comprehensive, Automatically Updated Fungal ITS Sequence Dataset for Reference-Based Chimera Control in Environmental Sequencing Efforts. *Microbes and Environments* advpub. doi: 10.1264/j sme2.ME14121

- Nilsson RH, Veldre V, Hartmann M, Unterseher M, Amend A, Bergsten J, Kristiansson E, Ryberg M, Jumpponen A, Abarenkov K (2010) An open source software package for automated extraction of ITS1 and ITS2 from fungal sequences for use in high-throughput community assays and molecular ecology. *Fungal Ecology* 3: 284–287. doi: 10.1016/j.funeco.2010.05.002
- O'Brian HE, Parrent JL, Jackson JA, Moncalvo J-M, Vilgalys R (2005) Fungal community analysis by large scale sequencing of environmental samples. *Applied and Environmental Microbiology* 71: 5544–5550. doi: 10.1128/AEM.71.9.5544-5550.2005
- Odelberg SJ, Weiss RB, Hata A, White R (1995) Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. *Nucleic Acids Research* 23: 2049–2057. doi: 10.1093/nar/23.11.2049
- Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J (2001a) Evaluation of PCR-Generated Chimeras, Mutations, and Heteroduplexes with 16S rRNA Gene-Based Cloning. *Appl Environ Microbiol* 67: 880–887. doi: 10.1128/aem.67.2.880-887.2001
- Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J (2001b) Evaluation of PCR-Generated Chimeras, Mutations, and Heteroduplexes with 16S rRNA Gene-Based Cloning. *Applied and Environmental Microbiology* 67: 880–887. doi: 10.1128/aem.67.2.880-887.2001
- Ryberg M, Nilsson RH, Kristiansson E, Töpel M, Jacobsson S, Larsson E (2008) Mining metadata from unidentified ITS sequences in GenBank: A case study in *Inocybe* (Basidiomycota). *BMC Evolutionary Biology* 8: 50. doi: 10.1186/1471-2148-8-50
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Consortium FB (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. doi: 10.1073/pnas.1117018109
- Smyth RP, Schlub TE, Grimm A, Venturi V, Chopra A, Mallal S, Davenport MP, Mak J (2010) Reducing chimera formation during PCR amplification to ensure accurate genotyping. *Gene* 469: 45–51. doi: 10.1016/j.gene.2010.08.009
- Tedersoo L, Bahram M, Dickie IA (2014) Does host plant richness explain diversity of ectomycorrhizal fungi? Re-evaluation of Gao et al. (2013) data sets reveals sampling effects. *Molecular Ecology* 23: 992–995. doi: 10.1111/mec.12660
- Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. *Nucleic Acids Research* 30: 2083–2088. doi: 10.1093/nar/30.9.2083
- 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.
- Wang GC, Wang Y (1997) Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Applied and Environmental Microbiology* 63: 4645–4650.
- White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (Eds) *PCR Protocols, A Guide to Methods and Applications*. Academic Press, San Diego, 315–322.

- Wright ES, Yilmaz LS, Noguera DR (2012) DECIPHER, a Search-Based Approach to Chimera Identification for 16S rRNA Sequences. *Applied and Environmental Microbiology* 78: 717–725. doi: 10.1128/aem.06516-11
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31: 3406–3415. doi: 10.1093/nar/gkg595

Distributional records of Antarctic fungi based on strains preserved in the Culture Collection of Fungi from Extreme Environments (CCFEE) Mycological Section associated with the Italian National Antarctic Museum (MNA)

Laura Selbmann⁴, Silvano Onofri⁴, Laura Zucconi⁴, Daniela Isola²,
Marino Rottigni^{1,2}, Claudio Ghiglione^{1,3}, Paola Piazza^{1,2},
Maria Chiara Alvaro^{1,2}, Stefano Schiaparelli^{1,2}

1 Italian National Antarctic Museum (MNA, Section of Genoa), University of Genoa, Genoa, Italy **2** Department of Earth, Environmental and Life Sciences (DISTAV), University of Genoa, Genoa, Italy **3** Department of Physical, Earth and Environmental Sciences, University of Siena, Siena, Italy **4** Department of Ecological and Biological Sciences (DEB), Tuscia University, Viterbo, Italy

Corresponding author: *Stefano Schiaparelli* (stefano.schiaparelli@unige.it)

Academic editor: *C.L. Schoch* | Received 21 May 2015 | Accepted 26 June 2015 | Published 16 July 2015

Citation: Selbmann L, Onofri S, Zucconi L, Isola D, Rottigni M, Ghiglione C, Piazza P, Alvaro MC, Schiaparelli S (2015) Distributional records of Antarctic fungi based on strains preserved in the Culture Collection of Fungi from Extreme Environments (CCFEE) Mycological Section associated with the Italian National Antarctic Museum (MNA). MycoKeys 10: 57–71. doi: 10.3897/mycokeys.10.5343

Resource citation: Italian National Antarctic Museum (2015-). Distributional records of Antarctic Fungi based on strains preserved in the Culture Collection of Fungi from Extreme Environments (CCFEE), Mycological Section associated with the Italian National Antarctic Museum (MNA). Contributed by: Laura Selbmann, Silvano Onofri, Laura Zucconi, Daniela Isola, Marino Rottigni, Claudio Ghiglione, Paola Piazza, Maria Chiara Alvaro, Stefano Schiaparelli. Online at <http://ipt.biodiversity.aq/manage/resource-register> Resource.do. GBIF key: <http://www.gbif.org/dataset/65e34334-66b3-4536-ac3c-857893fe0f6a>

Abstract

This dataset includes information regarding fungal strains collected during several Antarctic expeditions: the Italian National Antarctic Research program (PNRA) expeditions “X” (1994/1995), “XII” (1996/1997), “XVII” (2001/2002), “XIX” (2003/2004), “XXVI” (2010/2011), the Czech “IPY Expedition” (2007–2009) and a number of strains donated by E. Imre Friedmann (Florida State University) in

2001, isolated from samples collected during the U.S.A. Antarctic Expeditions of 1980-1982. Samples, consisting of colonized rocks, mosses, lichens, sediments and soils, were collected in Southern and Northern Victoria Land of the continental Antarctica and in the Antarctic Peninsula. A total of 259 different strains were isolated, belonging to 32 genera and 38 species, out of which 12 represented new taxa. These strains are preserved in the Antarctic section of the Culture Collection of Fungi from Extreme Environments (CCFEE), which represents one of the collections associated with the Italian National Antarctic Museum (MNA, Section of Genoa, Italy), located at the Laboratory of Systematic Botany and Mycology, Department of Ecological and Biological Sciences (DEB), Tuscia University (Viterbo, Italy). The CCFEE hosts a total of 486 Antarctic fungal strains from worldwide extreme environments. Distributional records are reported here for 259 of these strains. The holotypes of the 12 new species included in this dataset are maintained at CCFEE and in other international collections: CBS-KNAW Fungal Biodiversity Centre (Utrecht, Netherlands); DBVPG, Industrial Yeasts Collection (University of Perugia, Italy); DSMZ, German Collection of Microorganisms and Cell Cultures (Brunswick, Germany); IMI, International Mycological Institute (London, U.K.).

Key words

Antarctica, Antarctic Peninsula, CCFEE, Fungi, MNA, Victoria Land

Purpose

The aim of this study is to provide new distributional data on Antarctic fungi collected in the framework of past and recent expeditions (Italian PNRA, USA and Czech expeditions) and now preserved in the Antarctic section of the Culture Collection of Fungi From Extreme Environments (CCFEE - Antarctic Fungi), a collection associated with the Italian Antarctic National Museum (MNA, Section of Genoa, Italy). CCFEE is located at the Laboratory of Systematic Botany and Mycology, Department of Ecological and Biological Sciences (DEB) of the Tuscia University (Viterbo, Italy).

The dataset is the third Italian contribution to ANTABIF based on materials stored at the Italian National Antarctic Museum (MNA) and in its associated collections. The first MNA dataset published regarded the distributional records of Antarctic Mollusca, collected in the framework of the Latitudinal Gradient Program (Ghiglione et al. 2013), while the second MNA dataset regarded the distributional records of Antarctic Tanaidaceans collected in the Ross Sea (Piazza et al. 2014).

Project details

Project title: Antarctic Fungi from museum samples preserved in the Culture Collection of Fungi From Extreme Environments (CCFEE - Antarctic Fungi), which is a collection associated with the Italian National Antarctic Museum (MNA, Section of Genoa, Italy) hosted in the Laboratory of Systematic Botany and Mycology, Dept of Ecological and Biological Sciences (DEB), Tuscia University (Viterbo, Italy).

Curator and Promoter: Stefano Schiaparelli

Personnel: Laura Selbmann, Silvano Onofri, Laura Zucconi, Daniela Isola, Marino Rottigni, Claudio Ghiglione, Paola Piazza, Maria Chiara Alvaro, Stefano Schiaparelli.

Funding: The fungal strains were isolated from substrata (soils, sediments, mosses, lichens and rocks) sampled as part of different research projects funded by the Italian National Antarctic Research Program (PNRA), the National Science Foundation (NSF) of U.S.A. and Czech Ministry of Education, Youth and Sports, as listed below:

- PNRA X expedition (1994/1995) – Project 2d.1c (“Ecological and genetic aspects of population differentiation”);
- PNRA XII expedition (1996/1997) – Project 2c.3.1 (“Antarctic Biology: biodiversity, evolutionary strategies and colonization”);
- PNRA XVII expedition (2001/2002) – Project 2002/1.3 (“Antarctic terrestrial and freshwater ecosystems: response to latitudinal variations of climatic and environmental conditions”);
- PNRA XIX expedition (2003/2004) – Project 2002/1.3 (“Antarctic terrestrial and freshwater ecosystems: response to latitudinal variations of climatic and environmental conditions”);
- PNRA XXVI expedition (2010/2011) – Project 2009/A1.11 (“Biodiversity and evolution of polar rocks’ meristematic fungi: potential bioindicators of global changes and anthropogenic impacts”);
- Czech IPY expedition (2007-2009) – Projects ME-945 KONTAKT and MSM 0021622416 (“Multidisciplinary study of Antarctic terrestrial vegetation within the IPY framework”);

The collection also includes a number of strains donated by E. Imre Friedmann (Florida State University) in 2001, isolated from samples collected during the U.S.A. Antarctic Expeditions (1980-1982).

Study area description: Samples from which strains were isolated were collected in Southern and Northern Victoria Land, McMurdo Dry Valleys and the Antarctic Peninsula, from sea level up to 3400 m a.s.l. (Archambault Ridge, Antarctica).

Temporal coverage and coordinates:

PNRA X expedition (1994/1995): latitude from -76.91667 to -72.63333; longitude from 160.91667 to 169.36667

PNRA XII expedition (1996/1997): latitude from -74.06667 to -76.91775; longitude from 158.03517 to 165.31667

PNRA XVII expedition (2001/2002): latitude from -62.09139 to -74.88333; longitude from -58.45 to 163.71667

PNRA XIX expedition (2003/2004): latitude from -78.91075 to -72.71194; longitude from 159.04619 to 166.61667

PNRA XXVI expedition (2010/2011): latitude from -76.91000 to -73.44611; longitude from 159.18194 to 165.31667

U.S.A. expeditions (1980-1982): latitude from -77.86667 to -77.5; longitude from 160.66667 to 162

Czech expeditions (2007-2009): latitude -63.8; longitude -57.88333

Design description: Data were gathered by assembling distributional records of Antarctic fungal strains stored in the CCFEE. Samples from which all the strains were isolated were obtained in the framework of different past research expeditions, which had different aims and geographical targets. The main purpose of PNRA “X” expedition was to evaluate the adaptation of microfungal strains to low temperatures, thermal stress, desiccation and exposure to radiation. During the “XII”, “XVII” and “XIX” PNRA expeditions, the projects aimed at collecting rock samples for studying the biology and biodiversity of fungi from cryptoendolithic communities in the sandstone formations of several continental Antarctic coastal, and inland locations and from Antarctic Peninsula, as well as their role in the alteration of the rock substrate. The project developed in the framework of the “XXVI” PNRA expedition aimed at studying responses of rock meristematic fungi to environmental stresses related with the global change, and their role in the bioaccumulation and biomagnification processes of contaminants in the trophic networks. The main purpose of the Czech expedition was to study fungal biodiversity from soil and rock samples collected in the Antarctic Peninsula. The U.S.A. expeditions aimed at studying the biology and diversity of cryptoendolithic communities of the McMurdo Dry Valleys, and monitoring nano-climate variations within the rock substrate.

Methods

Method step description: See sampling description below and flowchart of Figure 1.

Study extent description: The distributional data herein considered refer to 53 different locations (Figure 2a, 2b, 2c), located along the Victoria Land (continental Antarctica) and the Antarctic Peninsula. Samples from which fungi were isolated were collected in the framework of the following expeditions, from 1980 to 2011:

- PNRA X expedition: samples of soil, mosses, lichens, rocks and sediments were collected from lake area near Camp OASI, Skua Lake, Tethys Bay, Enigma Lake, Vegetation Island, Inexpressible Island, Starr Nunatak, Edmonson Point, Baker Rocks, Kay Island, Mt Melbourne. In particular, in the site called Cryptogam Ridge, above the Mt. Melbourne volcanic cone, a small quantity of soil has been sampled for isolating thermophilic fungal strains and studying the fungal diversity of the area;
- PNRA XII expedition: samples of colonized sandstone were collected from Battleship Promontory, Mt Billing, Mt Howard, Trio Nunataks, Richards Nunatak, Timber Peak. Other sites have been visited to collect different types of rock samples as granites (BTN Base, Camp Antenne, Tethys Bay, Gondwana Station, O’Kane Glacier, Vegetation Island, Inexpressible Island, Boomerang Glacier, Collins Peak, Mt Crummer), pegmatite (Inexpressible Island, Boomerang Glacier), and sienite at Vegetation Island;

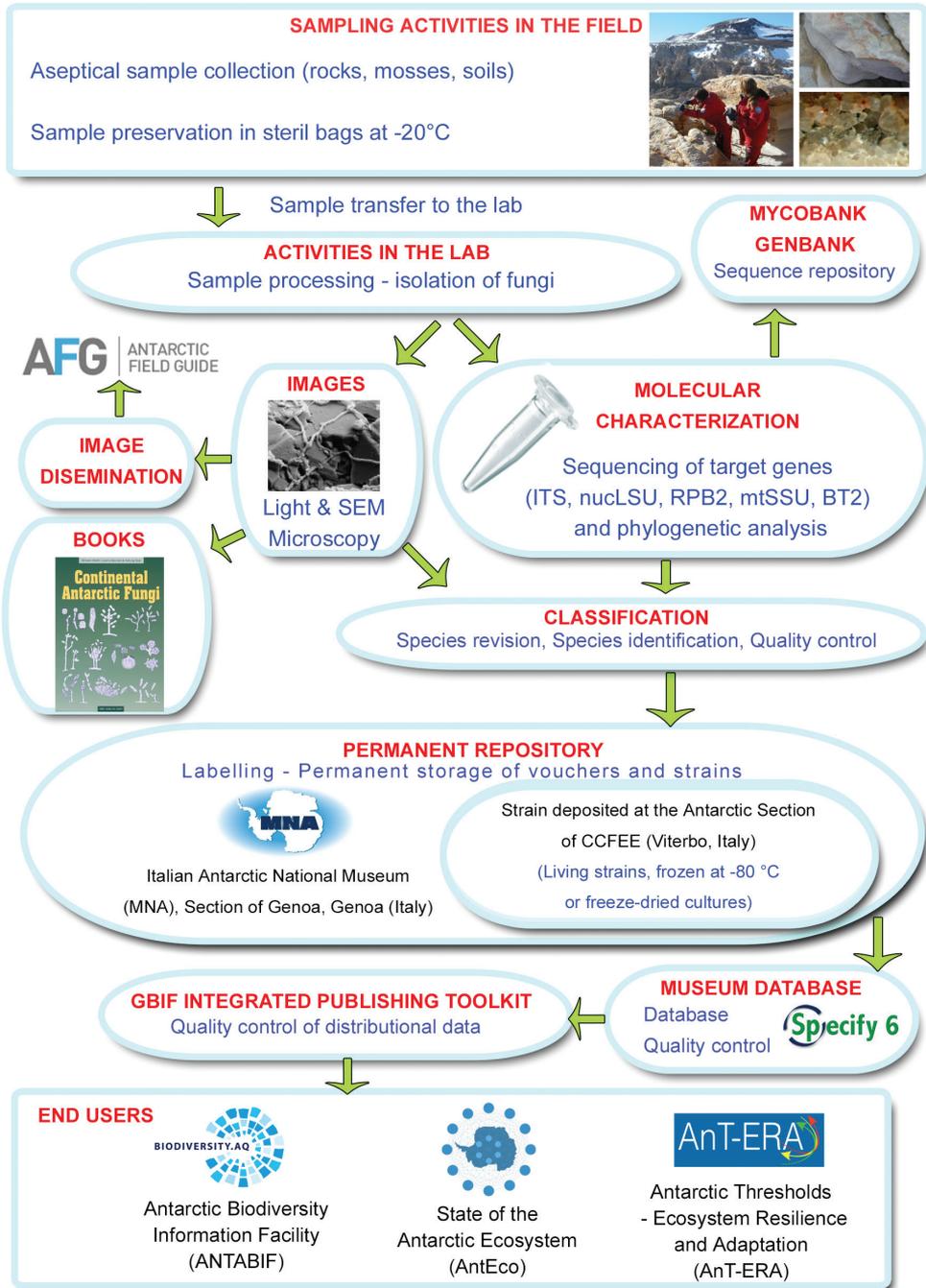


Figure 1. Flowchart depicting major steps in dataset development and publishing.

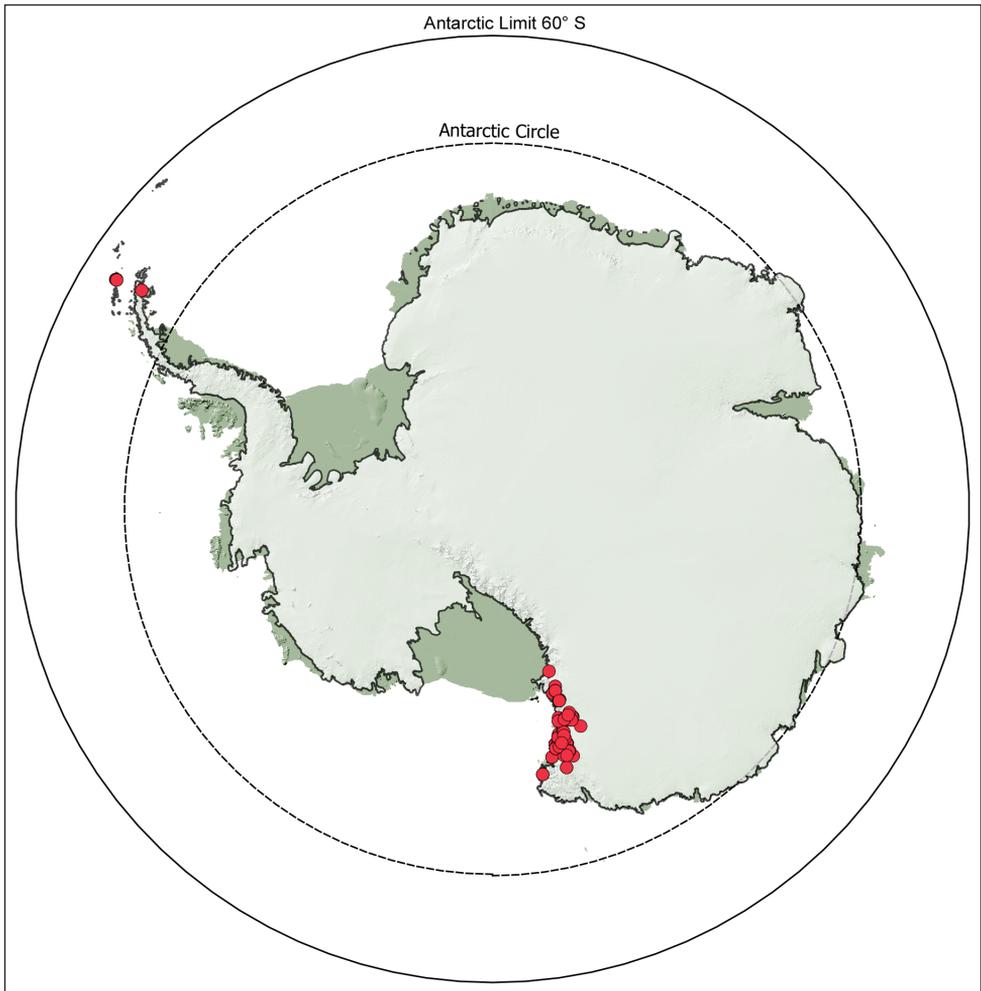


Figure 2a. Map of sampling stations.

- PNRA XVII expedition: different rock samples were collected in the Antarctic Peninsula starting from the Polish base Arctowski, King George Island;
- PNRA XIX expedition: sampling activities were carried out at Baker Rocks, Campo Icaro, Cape King, Edmonson Point, Ford Peak, Gondwana Lake, Harrow Peak, Inexpressible Island, Kay Island, Lichen Hills, Morris Basin (Ricker Hills), Starr Nunatak, Tarn Flat, Vegetation Island, and Widowmaker Pass in Northern Victoria Land, and Battleship Promontory, Convoy Range, Linnaeus Terrace, and University Valley in the McMurdo Dry Valleys, Southern Victoria Land;
- PNRA XXVI expedition: sampling was performed through latitudinal transect from 73°26'47"S, at Chisholm Hills (Cosmonaut Glacier), to 76°54'36"S, at Battleship Promontory, in a altitudinal gradient, from the sea level to 3400 m of Mount Adamson and, in a longitudinal gradient, from the coast to the inland;

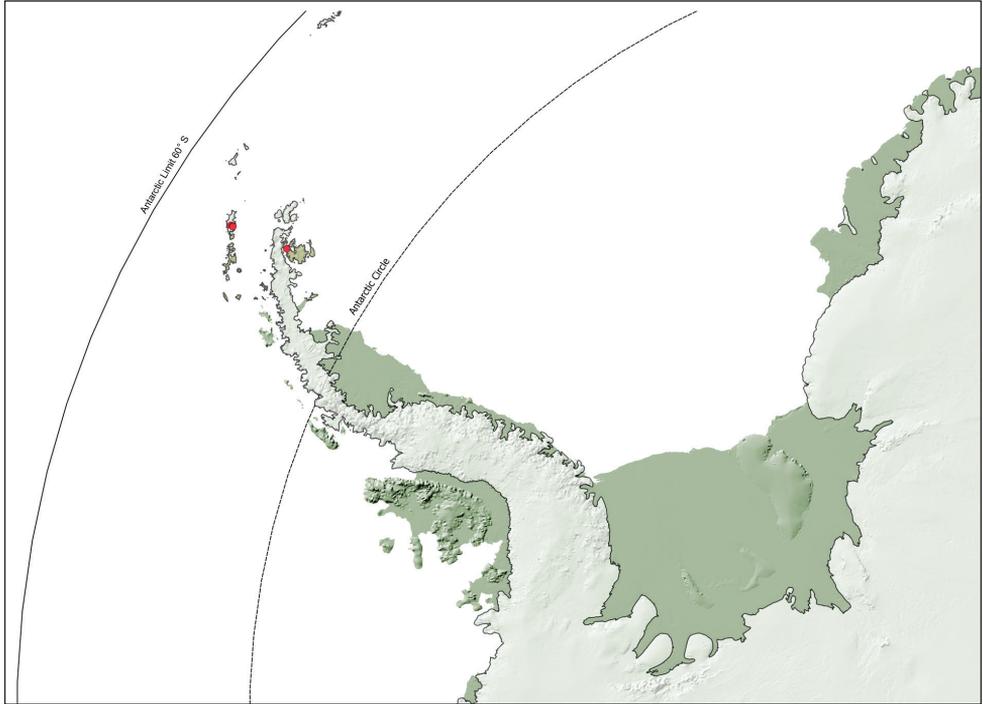


Figure 2b. Map of sampling stations. Zoom for the Antarctic Peninsula area.

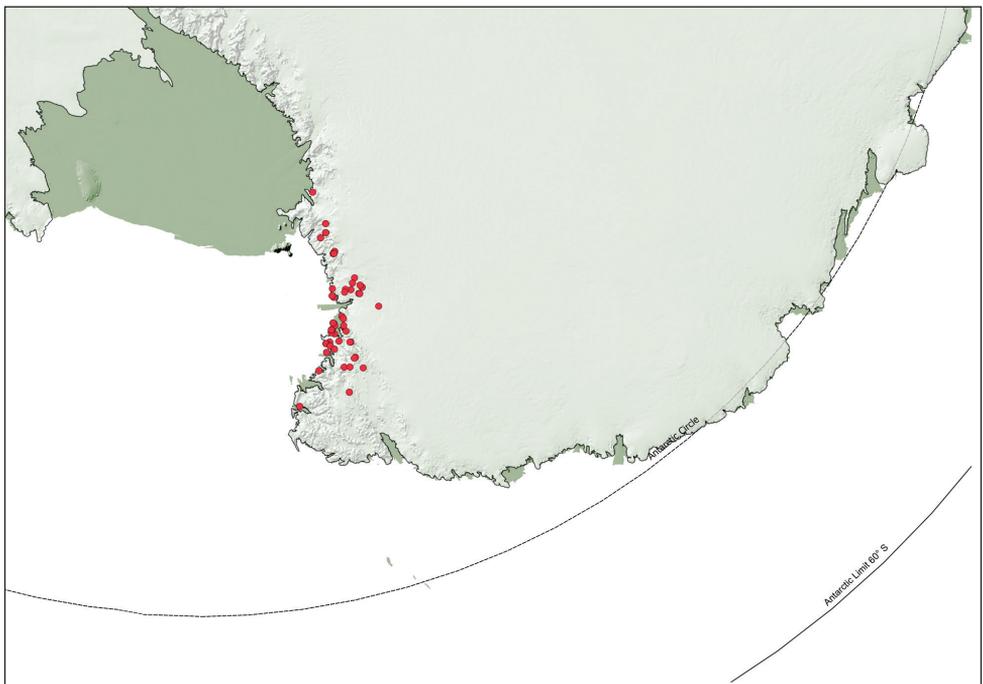


Figure 2c. Map of sampling stations. Zoom for the Ross Sea area.

all the sandstone's formations and some granite formations of Northern Victoria Land were visited;

- Czech IPY expedition: sampling was performed in the Antarctic Peninsula. Black meristematic fungi were isolated in Masaryk University, Brno, Czech Republic and donated to the CCFEE;

The U.S.A. expeditions: aimed to study the biology and diversity of cryptoendolithic communities of the McMurdo Dry Valleys, and sampling was performed in that area. Black meristematic fungi isolated were donated to the CCFEE.

Sampling description: The materials were collected by hand and, whenever needed, by using hammer and chisel in order to chop rock fragments from large boulders.. Samples were then placed in sterile plastic bags and stored at $-20\text{ }^{\circ}\text{C}$ (Figure 1). Fungi from rocks were isolated by crushing stones and plating fragments directly on Petri dishes containing Malt Agar (MA) amended with chloramphenicol 100 ppm; plates were incubated at $10\text{ }^{\circ}\text{C}$ and growth was inspected monthly.

Isolation from lichens was performed according to Selbmann et al. (2013). Isolation from other sediments, soils, and mosses was performed by diluting 1g of each substrate in 1 l of sterile water and plating 100 μl of the suspension in a 9 cm Petri dish containing MA or Czapek dox agar (CZA). Plates were incubated at 10 and $25\text{ }^{\circ}\text{C}$ and growth inspected every day. Colonies were transferred in fresh medium immediately after their appearance and identified by macro- and microscopic observations. Scanning electron microscopy observations were performed according to the methods described by Onofri et al. (1980). When morphology was not informative enough for identification, the isolates were studied with a molecular approach according to Selbmann et al. (2005; 2008). Fungi were identified by Selbmann, Onofri, Zucconi and Isola.

All the isolates are preserved in the Culture Collection of Fungi from Extreme Environments (CCFEE). All the fungi are preserved as living (at $4\text{ }^{\circ}\text{C}$), frozen (at $-80\text{ }^{\circ}\text{C}$) or freeze-dried cultures, or as dry samples. Holotype materials are deposited at: CBS-KNAW Fungal Biodiversity Centre (Utrecht, Netherlands); Industrial Yeasts Collection (Perugia, Italy); German Collection of Microorganisms and Cell Cultures (Brunswick, Germany); International Mycological Institute (London, United Kingdom) (Table.1). Ex-types of newly described taxa are preserved in the CCFEE.

The present dataset of Antarctic fungi has been formatted in order to fulfil Darwin Core standards required by the IPT scheme, according the GBIF Data Toolkit (<http://www.gbif.org/resources/2573>). The dataset was uploaded in the ANTABIF database (the geospatial component of GBIF).

Quality control description: Strains were identified by both morphological and molecular tools. DNA sequence data obtained from the studied strains are deposited in Mycobank and GenBank. Species names were crosschecked and made consistent with current ones as reported in MycoBank (last check made on 2015.08.04) (Figure 1). Distributional records' geographical coordinates are as published in the different papers reported in the reference section. During the more recent expeditions (XIX and XXVI) the coordinates were recorded using a Garmin GPS.

Table 1. List of Ex-types in the CCFEE associated collection of MNA with others collection codes referred to the holotypes' ones. Culture Collection of Fungi from Extreme Environments (Antarctic Fungi associated collection of MNA, Section of Genoa, University of Tuscia in Viterbo, Italy); **MNA** = Antarctic National Museum (MNA, Section of Genoa, Italy); **CBS** = Centraalbureau voor Schimmelcultures (Fungal Biodiversity Centre in Utrecht, Netherlands); **DBVPG** = Industrial Yeasts Collection (University of Perugia, Italy); **DSMZ** = German Collection of Microorganisms and Cell Cultures (Brunswick, Germany); **IMI** = International Mycological Institute (London, U.K). References: **1** = Egididi et al. 2014; **2** = Onofri et al. 1999; **3** = Selbmann et al. 2005; **4** = Selbmann et al. 2008; **5** = Selbmann et al. 2014.

Species	CCFEE	MNA	CBS	DBVPG	DSMZ	IMI	Gene	GeneBankAccession#	Ref.
<i>Vermiconia antarctica</i> Egididi & Selbmann	5488	F-0408	136107				ITS; nucLSU; RPB2	KF309982; KF309982; KF310089	1
<i>Tapprina antarctica</i> Selbmann & Turchetti	5198	F-0212	13532	5268	27485		ITS; nucLSU	JX124723; JX124717	5
<i>Rachicladosporium mcmurdoi</i> Selbmann & Onofri	5211	F-0225	119432				ITS; nucLSU; RPB2; BT2	KF309938; GU250371; KF310078; KF546731	1
<i>Rachicladosporium antarcticum</i> Onofri & Egididi	5527	F-0421					ITS	KF309942	1
<i>Oleogutnula mirabilis</i> Selbmann & de Hoog	5523	F-0417	136102				ITS; nucLSU	KF309973; KF310031	1
<i>Friedmanniomyces simplex</i> Selbmann, de Hoog, Mazzaglia, Friedmann & Onofri	5184	F-0198	116775				ITS; nucLSU; SSU; mtSSU	DQ028271; GU250368; DQ066716; GU250416	3
<i>Friedmanniomyces endolithicus</i> Onofri	5208	F-0222				379654	ITS	JN885549	2
<i>Extremus antarcticus</i> Selbmann & de Hoog	451	F-0001	136103				ITS; nucLSU; SSU; mtSSU; RPB2; BT2	KF309978; GU250360; GU250314; GU250403; KF310085; KF546729	1
<i>Elastiomyces elasticus</i> Zuccconi & Selbmann	5313	F-0325	122538				ITS	FJ415474	4
<i>Cryomyces minteri</i> Selbmann, de Hoog, Mazzaglia, Friedmann & Onofri	5187	F-0201	116302				ITS; nucLSU; SSU; mtSSU	DQ028270; KC315859; KC315858; KC315880	3
<i>Cryomyces antarcticus</i> Selbmann, de Hoog, Mazzaglia, Friedmann & Onofri	534	F-0013	116301				ITS; SSU; mtSSU	DQ028269; DQ066713; GU250410	3

Taxonomic coverage

General taxonomic coverage description: Fungal assemblages in Antarctic rocks have been extensively studied with particular emphasis for free-living but also symbiotic fungal populations, which constitute an important part of epilithic and endolithic communities (Selbmann et al. 2014). The interest for these communities is strongly connected to their high adaptability to the Antarctic ice-free environmental conditions, a combination of extremely low temperatures, low water activity, oligotrophy and high solar irradiation; for these reasons they constitute a useful model for studying evolution and adaptations to extreme environments. On the whole, this dataset takes into account 259 fungal strains, out of the 486 available at CCFEE. This initial dataset corresponds to the species that have been well characterized from a taxonomic point of view and have all the metadata required by Darwin Core for publication as distributional records. They belong to three Divisions, i.e. Ascomycota, Basidiomycota and Mucoromycotina (Figure 3), corresponding to a total of 32 genera (Figure 4) and 38 species. Among them, the 93% belong to Ascomycota Phylum, that includes the almost totality of genera and the totality of new taxa present in the collection (12 species and 6 genera) (Table 1). One basidiomycetous new species, isolated from soil under moss in Kay Island, *Cryptococcus vaughanmartiniae* Turchetti, Blanchette & Arenz (CCFEE5495=DBVPG5862), was recently described (Turchetti et al. 2015).

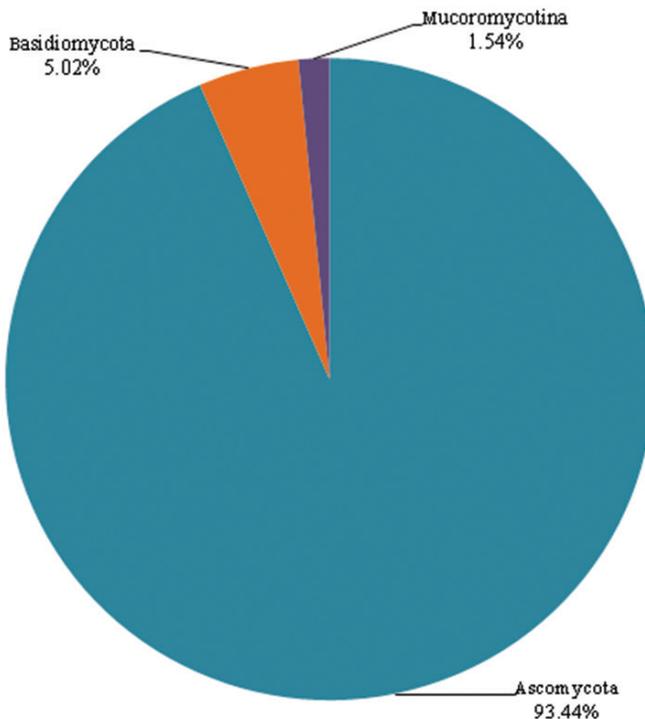


Figure 3. Taxonomic coverage (percentage per Division) of MNA Fungi Collection

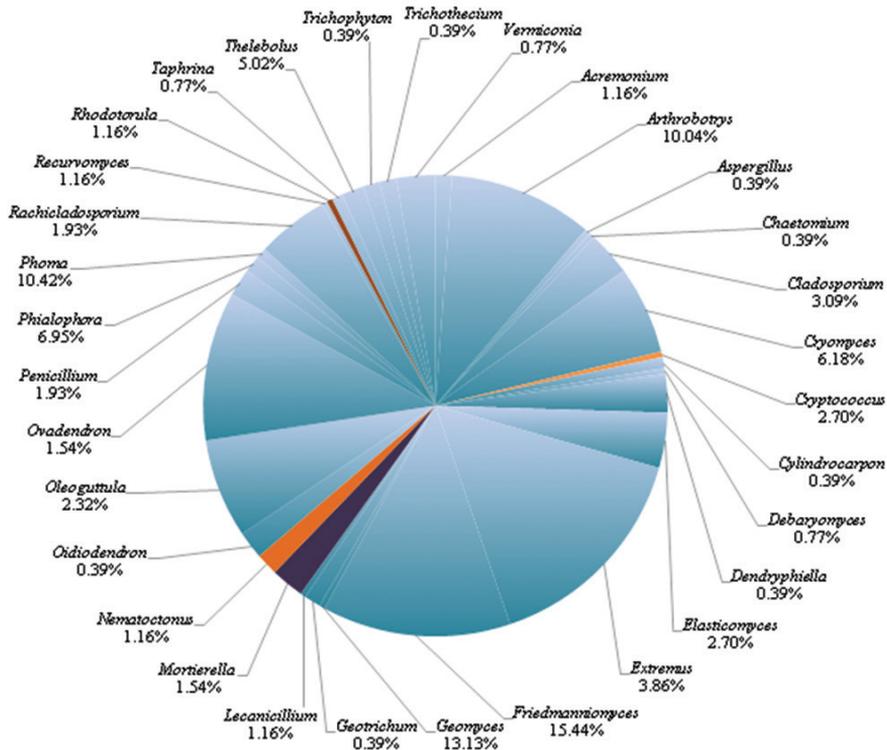


Figure 4. Taxonomic coverage (percentage per Genus) of MNA Fungi Collection.

Taxonomic ranks

Kingdom: Fungi

Division: Ascomycota

Genus: *Acremonium*, *Arthrobotrys*, *Aspergillus*, *Chaetomium*, *Cladosporium*, *Cryomyces*, *Cylindrocarpon*, *Debaryomyces*, *Dendryphiella*, *Elasticomyces*, *Extremus*, *Friedmanniomyces*, *Geomyces*, *Geotrichum*, *Lecanicillium*, *Oidiodendron*, *Oleoguttula*, *Ovadendron*, *Penicillium*, *Phialophora*, *Phoma*, *Rachicladosporium*, *Recurvomyces*, *Taphrina*, *Thelebolus*, *Trichophyton*, *Trichothecium*, *Vermiconia*

Species: *Acremonium berkeleyanum*; *Arthrobotrys ferox*; *Arthrobotrys* sp.; *Aspergillus versicolor*; *Chaetomium* sp.; *Cladosporium cladosporioides*; *Cladosporium herbarum*; *Cladosporium* sp.; *Cryomyces antarcticus*; *Cryomyces minteri*; *Cylindrocarpon didymum*; *Debaryomyces hanseni*; *Dendryphiella salina*; *Elasticomyces elasticus*; *Extremus antarcticus*; *Friedmanniomyces endolithicus*; *Friedmanniomyces simplex*; *Geomyces pannorum*; *Geomyces* sp.; *Geotrichum candidum*; *Lecanicillium muscarium*; *Oidiodendron griseum*; *Oleoguttula mirabilis*; *Oleoguttula* sp.; *Ovadendron sulphureo-ochraceum*; *Penicillium* sp.; *Penicillium viridicatum*; *Phialophora fastigiata*; *Phialophora malorum*; *Phialophora* sp.; *Phoma herbarum*; *Phoma* sp.; *Rachicladosporium antarcticum*; *Rachicladosporium mcmurdoi*; *Rachicladosporium* sp.; *Recurvomyces*

mirabilis; *Taphrina antarctica*; *Thelebolus microsporus*; *Trichophyton terrestre*; *Trichothecium roseum*; *Vermiconia antarctica*

Division: Basidiomycota

Genus: *Cryptococcus*, *Nematoctonus*, *Rhodotorula*

Species: *Cryptococcus antarcticus*; *Cryptococcus friedmannii*; *Cryptococcus victoriae*; *Cryptococcus vishniacii*; *Nematoctonus* sp.; *Rhodotorula mucilaginoso*; *Rhodotorula* sp.

Division: Mucoromycotina

Genus: *Mortierella*

Species: *Mortierella alpina*; *Mortierella* sp.

Natural collections description

Parent collection identifier: Italian National Antarctic Museum (MNA Section of Genoa, Italy), and Antarctic Section of the Culture Collection of Fungi from Extreme Environments (CCFEE), Mycological Section of MNA, Laboratory of Systematic Botany and Mycology, Dept of Ecological and Biological Sciences (DEB), Tuscia University (Viterbo, Italy).

Collection name: MNA (Section of Genoa) associated collection of Culture Collection of Fungi From Extreme Environments (CCFEE) - Antarctic Fungi

Collection identifier: <http://www.mna.it>; <http://www.sma.unitus.it/>

Specimen preservation method: Strains are stored and maintained as living cultures on Agar slant at 4 °C, frozen at -80 °C, freeze dried, or as dry samples.

Datasets

Dataset description: This dataset contains data from the Kingdom Fungi, Division Ascomycota, Basidiomycota and Mucoromycotina, from Victoria Land. In total it includes 38 different species of fungi for a total of 259 individuals. The Darwin Core elements included in the dataset are: catalogue number (i.e. MNA catalogue number), scientific name, expedition/station, gear, date collected, decimal latitude, decimal longitude, depth, individual count, institution code (i.e. name of institution where the samples are kept), collection code (i.e. MNA acronym) and status.

Object name: MNA (Section of Genoa) associated collection of CCFEE (Tuscia University) - Antarctic Fungi

Character encoding: UTF-8

Format name: Darwin Core Archive format

Format version: 1.0

Distribution: <http://ipt.biodiversity.aq/resource.do?r=aaa>

Language: English

Metadata language: English

License of use: This dataset [MNA (Section of Genoa) associated collection of CCFEE (Tuscia University) - Antarctic Fungi] is made available under the Open Data Commons Attribution License: <http://www.opendatacommons.org/license/by/1.0>

Date of metadata creation: April 8, 2015

Hierarchy level: Dataset

Acknowledgements

We are indebted to Cecile Gueidan (Australian National Herbarium) for the comments on a previous version of the MS.

References

References cited in the text

- Ghiglione C, Alvaro MC, Griffiths HJ, Linse K, Schiaparelli S (2013) Ross Sea Mollusca from the Latitudinal Gradient Program: R/V *Italia* 2004 Rauschert dredge samples. *ZooKeys* 341: 37–48. doi: 10.3897/zookeys.341.6031
- Onofri S, Castagnola M, Rossi Espagnet S (1980) Use of Scanning Electron Microscopy (SEM) in mycology. *Micologia Italiana* 1: 29–32.
- Piazza P, Blazewicz-Paszkowycz M, Ghiglione C, Alvaro MC, Schnabel K, Schiaparelli S (2014) Distributional records of Ross Sea (Antarctica) Tanaidacea from museum samples stored in the collections of the Italian National Antarctic Museum (MNA) and the New Zealand National Institute of Water and Atmospheric Research (NIWA). *ZooKeys* 451: 49–60. doi: 10.3897/zookeys.451.8373
- Selbmann L, Zucconi L, Onofri S, Cecchini C, Isola D, Turchetti B, Buzzini P (2014) Taxonomic and phenotypic characterization of yeasts isolated from worldwide cold rock-associated habitats. *Fungal Biology* 118: 61–71. doi: 10.1016/j.funbio.2013.11.002
- Selbmann L, Grube M, Onofri S, Isola D, Zucconi L (2013) Antarctic epilithic lichens as niches for black meristematic fungi. *Biology* 2: 784–797. doi: 10.3390/biology2020784
- Selbmann L, de Hoog GS, Zucconi L, Isola D, Ruisi S, den Ende AHGG, Ruibal C, De Leo F, Urzi C, Onofri S (2008) Drought meets acid: three new genera in a dothidealean clade of extremotolerant fungi. *Studies in Mycology*: 1–20. doi: 10.3114/sim.2008.61.01
- Selbmann L, de Hoog GS, Mazzaglia A, Friedmann EI, Onofri S (2005) Fungi at the edge of life: cryptoendolithic black fungi from Antarctic desert. *Studies in Mycology*: 1–32.
- Turchetti B, Selbmann L, Blanchette RA, Di Mauro S, Marchegiani E, Zucconi L, Arenz BE, Buzzini P (2015). *Cryptococcus vaughanmartinae* sp. nov. and *Cryptococcus onofrii* sp. nov.: two new species isolated from worldwide cold environments. *Extremophiles*, 19: 149–159. doi: 10.1007/s00792-014-0692-3

References based on this dataset

- de Vera JP, Boettger U, Noetzel RD, Sanchez FJ, Grunow D, Schmitz N, Lange C, Hubers HW, Billi D, Baque M, Rettberg P, Rabbow E, Reitz G, Berger T, Moller R, Bohmeier

- M, Horneck G, Westall F, Janchen J, Fritz J, Meyer C, Onofri S, Selbmann L, Zucconi L, Kozyrovska N, Leya T, Foing B, Demets R, Cockell CS, Bryce C, Wagner D, Serrano P, Edwards HGM, Joshi J, Huwe B, Ehrenfreund P, Elsaesser A, Ott S, Meessen J, Feyh N, Szewzyk U, Jaumann R, Spohn T (2012) Supporting Mars exploration: BIOMEX in Low Earth Orbit and further astrobiological studies on the Moon using Raman and PanCam technology. *Planetary and Space Science* 74: 103–110. doi: 10.1016/j.pss.2012.06.010
- Egidi E, de Hoog GS, Isola D, Onofri S, Quaedvlieg W, de Vries M, Verkley GJM, Stielow JB, Zucconi L, Selbmann L (2014) Phylogeny and taxonomy of meristematic rock-inhabiting black fungi in the Dothideomycetes based on multi-locus phylogenies. *Fungal Diversity* 65: 127–165. doi: 10.1007/s13225-013-0277-y
- Hubka V, Reblova M, Rehulka J, Selbmann L, Isola D, de Hoog SG, Kolarik M (2014) *Bradomyces* gen. nov. (Chaetothyriales, Trichomeriaceae), a new ascomycete genus accommodating poorly differentiated melanized fungi. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* 106: 979–992. doi: 10.1007/s10482-014-0267-4
- Onofri S, de la Torre R, de Vera JP, Ott S, Zucconi L, Selbmann L, Scalzi G, Venkateswaran KJ, Rabbow E, Inigo FJS, Horneck G (2012) Survival of Rock-Colonizing Organisms After 1.5 Years in Outer Space. *Astrobiology* 12: 508+. doi: 10.1089/ast.2011.0736
- Onofri S, Selbmann L, Barreca D, Isola D, Zucconi L (2009) Fungal survival in space conditions: new insights for lithopanspermia. *Plant Biosystems* 143: S85–S87. doi: 10.3114/sim.2008.61.10
- Onofri S, Barreca D, Selbmann L, Isola D, Rabbow E, Horneck G, De Vera J, Hatton J, Zucconi L (2008) Resistance of Antarctic black fungi and cryptoendolithic communities to simulated space and Martian conditions. *Studies in Mycology* 61: 99–109.
- Onofri S, Zucconi L, Tosi S (2007) Continental antarctic fungi.
- Onofri S, Selbmann L, Zucconi L, Pagano S (2004) Antarctic microfungi as models for exobiology. *Planetary and Space Science* 52: 229–237. doi: 10.1016/j.pss.2003.08.019
- Onofri S, Fenice M, Cicalini AR, Tosi S, Magrino A, Pagano S, Selbmann L, Zucconi L, Vishniac HS, Ocampo-Friedmann R, Friedmann FI (2000) Ecology and biology of microfungi from Antarctic rocks and soils. *Italian journal of Zoology* 67: 163–167. doi: 10.1080/11250000009356372
- Onofri S, Pagano S, Zucconi L, Tosi S (1999) *Friedmanniomyces endolithicus* (Fungi, Hyphomycetes), anam-gen and sp nov, from continental Antarctica. *Nova Hedwigia* 68: 175–182.
- Onofri S, Tosi S, Persiani AM, Maggi O, Riess S, Zucconi L (1992) Mycological researches in Victoria Land terrestrial ecosystems. In: *Proceedings of the 2nd Meeting on Biology in Antarctica*, Padua, 26–28.
- Ruibal C, Gueidan C, Selbmann L, Gorbushina AA, Crous PW, Groenewald JZ, Muggia L, Grube M, Isola D, Schoch CL, Staley JT, Lutzoni F, de Hoog GS (2009) Phylogeny of rock-inhabiting fungi related to Dothideomycetes. *Studies in Mycology* 64: 123–133. doi: 10.3114/sim.2009.64.06
- Scalzi G, Selbmann L, Zucconi L, Rabbow E, Horneck G, Albertano P, Onofri S (2012) LIFE Experiment: Isolation of Cryptoendolithic Organisms from Antarctic Colonized

- Sandstone Exposed to Space and Simulated Mars Conditions on the International Space Station. *Origins of Life and Evolution of Biospheres* 42: 253–262. doi: 10.1007/s11084-012-9282-5
- Selbmann L, Turchetti B, Yurkov A, Cecchini C, Zucconi L, Isola D, Buzzini P, Onofri S (2014) Description of *Taphrina antarctica* f.a. sp nov., a new anamorphic ascomycetous yeast species associated with Antarctic endolithic microbial communities and transfer of four *Lalaria* species in the genus *Taphrina*. *Extremophiles* 18: 707–721. doi: 10.1007/s00792-014-0651-z
- Selbmann L, Isola D, Fenice M, Zucconi L, Sterflinger K, Onofri S (2012) Potential extinction of Antarctic endemic fungal species as a consequence of global warming. *Science of the Total Environment* 438: 127–134. doi: 10.1016/j.scitotenv.2012.08.027
- Selbmann L, Isola D, Zucconi L, Onofri S (2011) Resistance to UV-B induced DNA damage in extreme-tolerant cryptoendolithic Antarctic fungi: detection by PCR assays. *Fungal Biology* 115: 937–944. doi: 10.1016/j.funbio.2011.02.016
- Selbmann L, Zucconi L, Ruisi S, Grube M, Cardinale M, Onofri S (2010) Culturable bacteria associated with Antarctic lichens: affiliation and psychrotolerance. *Polar Biology* 33: 71–83. doi: 10.1007/s00300-009-0686-2
- Tesei D, Marzban G, Zakharova K, Isola D, Selbmann L, Sterflinger K (2012) Alteration of protein patterns in black rock inhabiting fungi as a response to different temperatures. *Fungal Biol* 116: 932–940. doi: 10.1016/j.funbio.2012.06.004
- Zucconi L, Onofri S, Cecchini C, Isola D, Ripa C, Fenice M, Madonna S, Reboleiro-Rivas P, Selbmann L (2014) Mapping the lithic colonization at the boundaries of life in Northern Victoria Land, Antarctica. *Polar Biology*: 1–12. doi: 10.1007/s00300-014-1624-5
- Zucconi L, Selbmann L, Buzzini P, Turchetti B, Guglielmin M, Frisvad JC, Onofri S (2012) Searching for eukaryotic life preserved in Antarctic permafrost. *Polar Biology* 35: 749–757. doi: 10.1007/s00300-011-1119-6
- Zucconi L, Pagano S, Fenice M, Selbmann L, Tosi S, Onofri S (1996) Growth temperature preferences of fungal strains from Victoria Land, Antarctica. *Polar Biology* 16: 53–61. doi: 10.1007/Bf02388735

