MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2016) doi: 10.1111/1755-0998.12507

Don't put all your eggs in one basket: a cost-effective and powerful method to optimize primer choice for rRNA environmental community analyses using the Fluidigm Access Array

SHAWN P. BROWN,* ASTRID FERRER,* JAMES W. DALLING*† and KATY D. HEATH*
*Department of Plant Biology, University of Illinois at Urbana-Champaign, 505 S Goodwin Ave., Urbana, IL 61801, USA,
†Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Ancon, Panama

Abstract

With the increasing democratization of high-throughput sequencing (HTS) technologies, along with the concomitant increase in sequence yield per dollar, many researchers are exploring HTS for microbial community ecology. Many elements of experimental design can drastically affect the final observed community structure, notably the choice of primers for amplification prior to sequencing. Some targeted microbes can fail to amplify due to primer-targeted sequence divergence and be omitted from obtained sequences, leading to differences among primer pairs in the sequenced organisms even when targeting the same community. This potential source of taxonomic bias in HTS makes it prudent to investigate how primer choice will affect the sequenced community prior to investing in a costly community-wide sequencing effort. Here, we use Fluidigm's microfluidic Access Arrays (IFC) followed by Illumina® MiSeq Nano sequencing on a culture-derived local mock community to demonstrate how this approach allows for a low-cost combinatorial investigation of primer pairs and experimental samples (up to 48 primer pairs and 48 samples) to determine the most effective primers that maximize obtained communities whilst minimizing taxonomic biases.

Keywords: Fluidigm, fungi, internal transcribed spacer, large subunit, mock community, primer selection

Received 13 October 2015; revision received 19 January 2016; accepted 23 January 2016

Introduction

Microbial molecular ecology has undergone massive technological advancements in the last decade, allowing for deeper and ultra-high-throughput sequencing. This ability to directly interrogate microbial communities using locus-targeted sequencing of environmental DNA without relying on time-consuming and potentially biased culturing techniques has transformed our understanding of the hyperdiversity and basic ecology of microbial communities (Sogin *et al.* 2006; Jumpponen & Jones 2009). Although methodological obstacles in community sequencing are numerous including polymerase choice (Oliver *et al.* 2015), computational limitations (Schloss *et al.* 2009) and algorithm choice (Schmidt *et al.* 2015), one important aspect of experimental design remains insufficiently examined: the choice of the most efficient primers for amplification of environmental DNA.

Correspondence: Shawn P. Brown, Fax: 217-224-7246; E-mail: spbrown1@illinois.edu

Most often, researchers use oligonucleotide primers to amplify and sequence target regions of the genomes, focusing on regions that are both conserved enough to be recognized as orthologous, but also variable enough among species (and invariable enough within species) to be diagnostic to lower taxonomic ranks (*see* Lindahl *et al.* 2013). Ideal primers (*i.e.* universal primers) must reliably amplify all target taxa with similar efficiencies (*e.g.* 100% efficiency indicates absolute doubling of template DNA per cycle), enabling quantification of the microorganisms present in DNA extracted directly from the environment. Particularly challenging is the design of primers targeting fungi as fungi have diverged relatively recently from other Eukaryotes (*i.e.* they have similar target sequences), as compared to bacteria or archaea (Hugenholtz & Pace 1996); therefore, 'nontarget amplification' of other Eukaryotes during fungal amplification is common (Brown *et al.* 2015a,b). Primers also might introduce biases due to differential amplification efficiencies of different fungal groups (Bellemain *et al.* 2010; Ihrmark *et al.* 2012), a problem common to all environmental

sequencing. The preferential amplification of certain organismal groups over others in mixed environmental DNA produces an inferred community that is not representative of the environmental DNA template (Anderson & Cairney 2004), as final taxonomic distributions will be skewed towards the direction of the bias. Such bias has important implications for estimating total community diversity as well as abundances of particular taxonomic units.

Regions within the rRNA repeat are most often the targets for fungal community sequencing. The conventional wisdom is that the internal transcribed spacer (ITS) regions are superior targets for measurements of biodiversity, whereas targets on the large subunit (LSU) are superior if the aim of a study will incorporate phylogenetic measures such as queries on phylogenetic diversity. The ITS regions have been designated at the official fungal barcode (see Seifert 2009; Schoch & Seifert 2012; Schoch *et al.* 2012) to be used for taxonomic identification and phylogenetic placement of single organism samples. Consequently, ITS gene targets are well represented in the global genetic repositories of vouchered and cultured specimens, facilitating positive taxonomic identification of unknown samples (Kõljalg *et al.* 2013; Lindahl *et al.* 2013;. However, this barcoding initiative was not optimized to handle mixed environmental DNA. The ITS region for fungi is not amenable to global multiple sequence alignments at taxonomic ranks higher than Genus due to extreme interspecific sequence hypervariability (Nilsson *et al.* 2008, 2012; Kõljalg *et al.* 2013). Because of this, ITS regions prohibit robust integration of evolutionary hypotheses into a community analysis framework such as queries on community phylogenetic diversity (Brown & Jumpponen 2015). Given these limitations with ITS regions, LSU regions are gaining support for locus-targeted studies (Weber *et al.* 2013; Lothamer *et al.* 2014; Porras-Alfaro *et al.* 2014). The LSU is globally alignable and, moreover, the secondary structure afforded by multiple sequence alignments may support improved operational taxonomic unit (OTU) binning performance (Schloss 2013). Both ITS and LSU can successfully elucidate community ecological shifts in the queried community (Brown *et al.* 2014).

In addition to the difficult choice of ITS vs. LSU, many of the fungal primers commonly used for community sequencing were designed using a relatively small taxonomic pool and/or were chosen to maximize amplification of individual organismal samples rather than a mixed template. For example, the commonly used fungal primer ITS1f, which targets the internal transcribed spacer 1 (ITS1; Gardes & Bruns 1993), was designed using nine fungal sequences, six of which belonged to the same Order (Boletales). Primers targeting the ITS regions have been found to amplify some non-dikarya fungi poorly (Bellemain *et al.* 2010), including the Glomeromycota (Renker *et al.* 2003) and the zoosporic fungi (Marano *et al.* 2012); even when present in the environment, these taxa may be rarely detected in sequence-based ecology. The limited taxonomic scope in the design of common fungal universal primers has led to a concerted effort to design new primers that reduce taxonomic biases in next-generation sequencing studies (*see* Ihrmark *et al.* 2012; Toju *et al.* 2012; Mueller *et al.* 2016). These newer 'universal' primers might be better at capturing a more representative fungal community; nevertheless, primers cannot be tested against all fungal groups and still might suffer from unforeseen amplification biases.

One way to investigate the suitability of different primers prior to sequencing of environmental DNA is to apply a community sequencing framework to a mock community of 'local' taxa. That is, a mock community made up of isolates that are collected directly from and concurrently with samples collected for environmental microbial community analyses. The use of a mock community is not limited to fungal community analyses and is suitable for any targeted organismal groups for which individuals can be cultured. Many locus-targeted community sequencing efforts utilize a mock community (*see* Amend *et al.* 2010; Caporaso *et al.* 2011; Bokulich *et al.* 2013; Nguyen *et al.* 2015); however, these are generally included to optimize OTU clustering parameters rather than to inform primer selection. Previous attempts to utilize a mock community to inform primer decision have been conducted using *in silico* tests to predict how thoroughly a primer pair will amplify a targeted region (Bokulich & Mills 2013). Whilst *in silico* methods are a powerful way to provide insight on primer efficacy, they are subject to insufficient database representation in the global genetic repositories. Most mock communities to date have consisted of commercial cultures or cultures from taxa that have a large global genetic database presence (Bokulich *et al.* 2013). As these local organisms are isolated in parallel with environmental sampling, the cultures represent a subset of the potential OTUs captured through culture-independent analyses (Xu *et al.* 2012) and are likely to be affected by the same primer-induced amplification biases as the targeted environmental samples. Of note, culture-dependent and culture-independent methods do not generate completely overlapping taxa (*see* Vaz-Moreira *et al.* 2011; Singh *et al.* 2012; Jumpponen *et al.* 2015), and thus even a local mock community may not address bias in the uncultured community members. However, local mock communities might have particular power to detect community-wide primer biases that would otherwise go undetected and to facilitate an informed decision on primer selection for maximum capture of targeted microbes.

The Fluidigm Access Array system (www.fluidigm.com) utilizes integrated fluidic circuits (IFC) and microfluidic methods to automate amplicon production for next-generation sequencing by combining template DNA (48 sample capacity) with up to 48 different primer pairs whilst conducting parallel PCRs in a fully enclosed environment – thus also reducing potential laboratory contamination sources and/or human-induced reagent addition errors (*see* Carlsen *et al.* 2012; Weiss *et al.* 2014; Nguyen *et al.* 2015 *for discussions of possible human or laboratory contamination*). The Fluidigm Access Array was developed to facilitate single-cell isolation for genomic resequencing efforts, but allows for a cost-effective method to amplify different target regions from multiple samples in parallel. Among other uses, Fluidigm could be used to amplify multiple taxonomically informative gene regions from isolates to expedite downstream phylogenetic analyses (Ferrer, *in prep*).

Here, we demonstrate the use of Fluidigm technology in conjunction with Illumina MiSeq Nano sequencing as an efficient test of primer suitability for community sequencing. In this example, we focus on optimizing primer selection for fungal community analysis, but this approach is also relevant for communities of bacteria and archaea. We first generated community DNA from a mock community assembled from an extensive single spore isolate culture collection, then amplified this mock community with several ITS and LSU primer combinations to optimize primer selection in preparation for an upcoming full-scale interrogation of environmental samples. We then ask: 1) Which primer sets result in the most unbiased representation of the mock community's *taxonomic distribution,* and, 2) Which primer sets result in the best representation of the mock community's *richness*? Our study demonstrates the utility of pairing a carefully designed local mock community with empirical assessment of primer efficacy for troubleshooting environmental sequencing studies in community ecology.

Materials and methods

Sampling and sequencing of local mock community

We used fungal communities inhabiting submerged wood, which were collected as part of a larger experiment to assess both the assembly of microbial communities across a salinity gradient on Coiba Island, Coiba National Park, Panama (7°28′N, 81°47′W) (Brown *et al.*, *in prep*), and to describe and catalog fungal biodiversity through a massive culturing effort (Ferrer *et al.*, *in prep*). Briefly, replicate sections of branch wood from *Guazuma ulmifolia* Lam. trees (Family Malvaceae, commonly called West Indian Elm) of 30 cm length and 2 cm diameter were anchored to the bottom of three rivers across four salinity gradients from fresh water to open ocean and sampled by removing three replicates every *ca.* 3 months for total genomic extraction and moist-chamber incubation.

Wood samples were placed in moist chambers consisting of a plastic box layered at the base with sterile damp absorbent paper. Samples were examined for fruiting bodies after collection using a compound microscope. Cultures were obtained from single spores placed on antibiotic water agar (0.5 mg streptomycin sulphate, 0.5 mg penicillin G, 18 g agar, 1L distilled water). Germinating spores were transferred onto potato dextrose agar (PDA). Fungal mycelia were scraped from axenic PDA cultures and DNA was extracted (DNAeasy Plant Mini Kit; Qiagen, Valencia, CA, USA) as per the manufacturer's protocol. In all, genomic DNA was extracted from a total of 180 unique fungal isolates. To characterize our local mock community, isolated genomic DNA of each culture was amplified using the fungal specific primers ITS1f and LR3 (Vilgalys & Hester 1990; Gardes & Bruns 1993, respectively) to target both ITS regions and the first variable domains of the LSU (D1 and D2). A volume of 2 μL of DNA template and 2.5 μL each primer (1 μM) were combined into wells of PuReTaq Ready-To-Go PCR Beads (GE Healthcare Limited, Little Chalfont, UK), and sterile water was added to a total volume of 25 μL. Samples were amplified under the following conditions: 94 °C initial denaturization for 5 min followed by 30 cycles of 94 °C for 30 s, 50 °C annealing and 72 °C extension followed by an additional final extension time of 10 min at 72 °C. PCR products were subsequently cleaned (QIAquick PCR Purification Kit; Qiagen), sequenced on an AB 3730xl DNA analyser (Applied Biosystems, Foster City, CA, USA), and the resultant sequences were manually trimmed and corrected using the program SEQUENCHER (v5.1).

In all, 166 high-quality Sanger sequences were obtained for the mock community (the remaining 14 consistently failed to amplify and were thus omitted from further consideration; these likely result from primer mismatch indicating that, in environmental sequencing surveys, even culturable taxa might be missed). To taxonomically place these mock sequences, obtained sequences were queried against only typified sequences in GenBank (BLASTn); placement along with BLAST score statistics is available in Table S1 (Supporting information). It is important to note that many of these mock community members are likely novel (species descriptions are ongoing), so the taxonomic placement of these mock community members at lower taxonomic ranks (genus, species) is likely to be skewed due to their relatively poor representation in the global genetic repositories of across kingdom Fungi. Familial and Ordinal level identifications are likely to be correct, and thus, we focus

on those. All mock sequences are available in the supplemental information and in GenBank (Accession nos KU535697-KU535859; DRYAD doi:10.5061/dryad.tn6vn).

Sequencing of the mock community with Fluidigm

A total of 1 μL of extracted genomic DNA from each mock isolate was homogenized into a single sample to act as a heterogeneous environmental template of the local mock community. We used one column of a microfluidic Fluidigm Access Array IFC chip (the remaining columns were used for unrelated samples) to combine our experimentally derived local mock community with all primer pair combinations along with all necessary PCR reagents. We used the default concentrations and PCR parameters native to Fluidigm (see www.fluidigm.com/documents for details). Amplicons were generated in 30 nL reactions using a two-step PCR protocol (total of 35 PCR cycles: 60 °C annealing temperature and 72 °C extension temperature). Primary PCR using native Fluidigm parameters amplified target DNA region using sample-specific primers in addition to Fluidigm-specific amplification primer pads (CS1 and CS2), resulting in targeted amplicons that consist of [CS1-5′ Forward PCR Primer-Amplicon-3′ Reverse PCR Primer-CS2]. Secondary PCRs added Illumina-specific sequencing linkers and sample-specific molecular identification tags (MID, 10 bp) with the final amplicon construct consisting of [Linker P5-CS1-5′Primer-Amplicon-3′Primer-CS2-MID-Linker P7]. Final amplicon constructs were quantified (Quant-iT PicoGreen dsDNA Assay Kit; Invitrogen, Carlsbad CA, USA) and combined to an equamolar concentration then sequenced bidirectionally on one ILLUMINA MISEQ NANO CHIP (V.2 chemistry using paired-end sequencing with 2 × 250 flows). Fluidigm amplification and sequencing was conducted at the W.M. Keck Center (Urbana, IL, USA).

Primers tested

Primer choice focused on the two most commonly used gene regions for fungal community analyses – internal transcribed spacer (ITS) and large subunit (LSU) of the ribosomal gene repeat. Primers selected here (see Table 1 for primer sequences melting temperatures and references) include three ITS primer pairs [(ITS1f-ITS2; *ITS1*), (fITS7-ITS4; *ITS2*), (ITS3KYO2-ITS4KYO3; *ITS2*)] and two LSU primer pairs [(LR0R-LR3; *D1 and D2*), (LR0R (Rehner)-nuLSU401; *D1 and D2*), in which the primer combination is followed by the targeted divergent regions in italics. It is important to note that each primer pair tested has different inherent melting temperatures and may thus respond differently during the annealing step when reactions are conducted at a consistent temperature.

Sequence analyses, bioinformatics and statistics

Sequences were processed using the program MOTHUR (v.1.33.3; Schloss *et al*. 2009). The obtained raw sequences were parsed into a forward and reverse fastq file that contained all sequences from the mock community (i.e. containing sequences for every primer pair used). These fastq files were contiged and screened for a minimum of 25-bp overlap of the paired reads, and the resultant file was used to parse these into a fasta file for each primer pair and each sample (five individual fasta files) based on the inclusion of forward and reverse primer sequence (1-bp variation was permissible for inclusion; see Table 1). For two primer pairs, failure to retain at least 70% of initial sequences after contiging was observed due to long amplicon length; these were reanalysed to include only the forward reads for all subsequent analyses [LR0R-LR3 and LR0R(Rehner)-nuLSU401]. These forward reads include full-length variable region D1 and partial D2 on the LSU. The ITS sequences were truncated

Table 1 List of ITS and LSU targeting primer pairs used to query mock communities along with the complete primer sequence, melting temperature and targeted divergent domains. Primer names and sequences in bold represent the forward sequencing primer for each pair

Targeted region	Primer pair	Primer sequence	Melting temperature	Targeted divergent domain(s)	References
ITS	**ITS1f**	**CTTGGTCATTTAGAGGAAGTAA**	49.7°	ITS1	Gardes & Bruns (1993)
	ITS2	GCTGCGTTCTTCATCGATGC	57.0°	ITS1	White *et al*. (1990)
	fITS7	**GTGARTCATCGAATCTTTG**	47.3°	ITS2	Ihrmark *et al*. (2012)
	ITS4	TCCTCCGCTTATTGATATGC	52.1°	ITS2	White *et al*. (1990)
	ITS3KYO2	**GATGAAGAACGYAGYRAA**	48.4°	ITS2	Toju *et al*. (2012)
	ITS4KYO3	CTBTTVCCKCTTCACTCG	52.5°	ITS2	Toju *et al*. (2012)
LSU	**LR0R**	**CCGCTGAACTTAAGCATATCA**	52.7°	D1, D2	Vilgalys & Hester (1990)
	LR3	CCGTGTTTCAAGACGGG	53.6°	D1, D2	Vilgalys & Hester (1990)
	LR0R(Rehner)	**GTACCCGCTGAACTTAAGC**	53.3°	D1, D2	Rehner & Samuels (1994)
	nuLSU401	CCTTTCAACAATTTCACGT	48.7°	D1, D2	Dörning *et al*. (2000)

(200 bp) such that all sequences were of equal length to avoid inflating genetic distance calculation prior to implementation of clustering algorithms (Brown *et al.* 2013). A paired fastq file of raw sequence data has been deposited in the SRA at NCBI (Bioproject PRJNA307545, Run SSR3083895).

We aligned LSU sequences (separately for each primer pair) against a modified LSU alignment based on the Assembling the Fungal Tree of Life (AFTOL) LSU alignment (James *et al.* 2006) where nonfungal sequences were removed and remaining sequences were *de novo* aligned using MUSCLE (*see* Brown *et al.* 2014). We identified and culled potential chimeras using UCHIME (Edgar *et al.* 2011) as implemented in MOTHUR. To correct for potential sequencing errors, we then applied pseudo-single-linkage clustering (pre.cluster; Huse *et al.* 2010).

The truncated ITS sequences were pairwise-aligned (Needleman & Wunsch 1970) to create a genetic distance matrix (separately for each primer pair), whereas the LSU distance matrix was generated from the multiple sequence alignment. The secondary structure generated from aligned LSU sequences improves OTU binning efficiency (Schloss 2013), but as fungal ITS sequences cannot be properly globally aligned, pairwise alignment was used instead. The resultant distance matrices were used to cluster sequences into operational taxonomic units (OTUs) at a 97% similarity threshold using the robust average-neighbour method (UPGMA) as implemented in MOTHUR (Schmidt *et al.* 2015). OTUs that had query coverage value and/or identity values <95% when queried (BLASTn) against a custom database consisting of Sanger sequences of the imputed mock community were considered artefacts of an unknown origin and omitted from further analyses (sequencing of unrelated samples was conducted on the same chip, these artefacts could be derived from tag-switching (Carlsen *et al.* 2012) or some other unidentified migration of samples across the chip). An annotated list of commands implemented for these analyses is provided as a supplemental file available on DRYAD (doi:10.5061/dryad.tn6vn).

Exploration for systemic taxonomic biases of primer pairs

Representative sequences of each OTU for each primer pair (DYRAD doi:10.5061/dryad.tn6vn) were queried (BLASTn) against the mock community sequences. First, we used a chi-square test of independence to ask whether, on a class level, the proportions of identified classes from each primer pair were significantly different from that of the known mock community. To explore the potential for taxonomic bias on an Order and Family level (chi-square tests on the Order and Family could not be done due to insufficient expected values for each category), we generated collector's curves using 1000 randomizations in the program ESTIMATES (v.9.1.0; Colwell 2006), affording a visual estimation of the efficacy of each primer set to capture all the Orders and Families from the mock community. All statistics were conducted using a combination of MOTHUR and JMP (v11.0.0; The SAS Institute, Cary, NC, USA).

Assessing primer efficiencies in returning mock community richness

To assess the ability of primers to capture the maximum diversity of our mock community, we calculated two richness metrics for each primer pair: (i) observed OTU richness (S_{obs}) and (ii) bias-controlled Chao1 estimate of extrapolated minimum richness (estimated richness if every last OTU were able to be accounted for) where Chao1 = $S_{obs} + [(n_1(n_1\text{-}1))/(2(n_2 + 1))]$ where n_1 is the number of local singleton OTUs, and n_2 is the number of local doubleton OTUs (Chao 1987). Richness metrics were estimated iteratively (1000 runs) at a subsample (without replacement) depth of 500 sequences per iteration as implemented in MOTHUR. Separate one-way analyses of variance (ANOVAS) for ITS and LSU regions were used to determine whether primer selection had a significant effect on richness estimators.

Results

Composition of the mock community

Of the 166 fungi from the mock community for which we obtained Sanger sequences, all but four belonged to Phylum Ascomycota (the remaining were Basidiomycota). Of these, six Classes, 20 Orders, 35 Families and 53 Genera were represented. Complete taxonomic identifications along with BLASTn alignment scores against GenBank representatives are provided in Table S1 (Supporting information). As commonly observed in aquatic systems, the majority of taxa belonged to the ascomycetous Classes Sordariomycetes (67.6%) and Dothideomycetes (18.3%).

Systemic taxonomic biases of primer pairs

Pearson's chi-squared tests for independence indicated that there were no substantial differences between the Class level taxonomic distributions in the mock community vs. the OTUs obtained using community sequencing ($P > 0.4$ for all primer pairs; Fig. 1a). Thus at the Class level, each primer pair captured similar communities. Differences emerged, however, when primer pairs were queried at lower taxonomic ranks. Collector's curves on

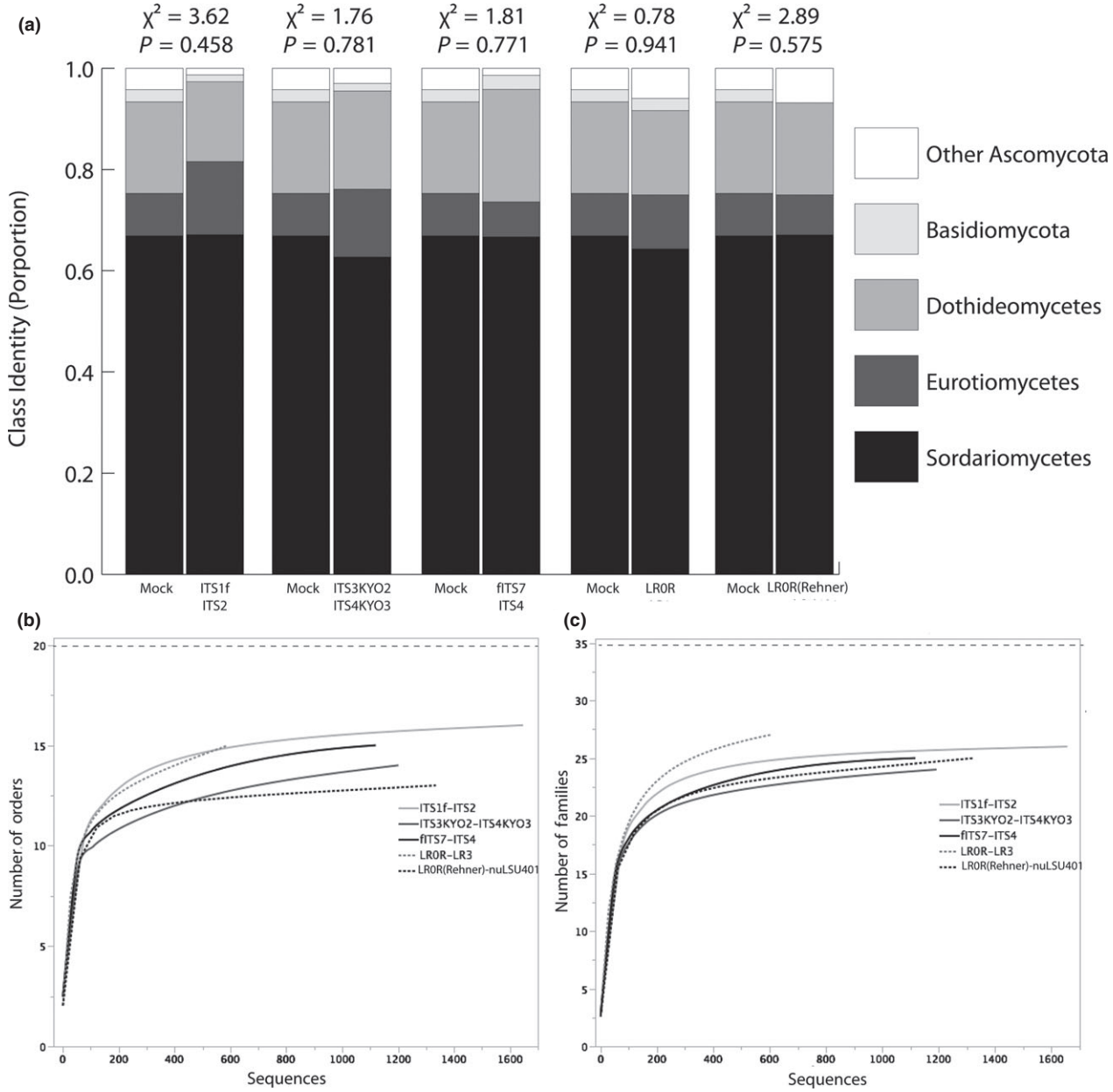

Fig. 1 The proportions of Classes captured by the five primer pair combinations are indistinguishable from the local mock community based on Pearson's chi-squared test statistics (a) and collector's curves on Order level identification (b) and Family level identification (c) of the five primer pairs suggest that at the Order and Family level, primer sets differentially capture the imputed mock community member. Horizontal dashed lines in (b) and (c) are the total number of Orders or Families in the Mock community. Consistently, primer set LR0R-LR3 captured more of available taxonomic diversity of the mock community.

the Order (Fig. 1b) and the Family (Fig. 1c) levels reveal differences among primer pairs in the amount of diversity captured. In general, the ITS primer pair ITS1f-ITS2 and the LSU primer pair LR0R-LR3 captured more orders and families than did the other primer pairs, with LR0R-LR3 outperforming all others. None of the primer pairs captured the single representative of the Order Ophiostomatales of the mock community; nor did they capture the rare members of the Families Anteagloniaceae, Corynesporascaceae, Cyphellophoraceae or Melanconidaceae – suggesting that these rare taxa might be discriminated against in mixed templates. Complete lists of OTUs and positive hits against the mock community are presented in Table S2 (Supporting information).

Assessing primer efficacy using a known mock community

None of the primer sets returned all of the pure culture species known to be included in the mock community. In fact, even the highest iterative observed richness (S_{obs}) estimates only returned about half of the known richness of the mock community – suggesting that all included primer pairs suffered from restricted taxon amplification, possibly due to the relatively shallow sequencing employed here. Nonetheless, the primer pair used to amplify the community had a significant effect on the estimated richness for both the ITS and LSU gene regions, indicating that choice of primer pair in our study would be expected to alter the snapshot of the fungal community acquired via amplicon sequencing.

For the ITS region, we found a highly significant, but biologically relatively minor, difference among observed richness estimates ($F_{2,2997}$ = 1573.98, P < 0.0001). Tukey's post hoc HSD indicated that all means were distinct from each other (Fig. 2a; Table S3, Supporting information), with the primer pair fITS7-ITS4 returning the highest richness (13.03% increase compared with the smallest estimate). Similar to observed richness estimators, we also found that the extrapolated minimum richness (Chao1) estimate depended on the ITS primer pair ($F_{2,2997}$ = 397.8, P < 0.0001), although the ranking of primers differed slightly from the observed richness, with the primer pair ITS1f-ITS2 returning a slightly higher Chao1 estimate (8.70% increase compared with the smallest estimate; Fig. 2b, Table S3, Supporting information).

For the LSU region, choice of primer pair had a much more pronounced effect on estimated community richness, compared with the ITS region. LSU primers resulted in significant differences in observed richness estimates (Fig. 2a), with the primer pair LR0R-LR3 returning the highest S_{obs} (24.67% increase compared with the smallest estimate, t = 123.21, P < 0.0001) and Chao1 (89.38% increase compared with the smallest estimate; t = 63.71, P < 0.0001) estimates**.** Importantly, the Chao1 estimated richness metric for the primer pair LR0R-LR3 (172.16 ± 35.66) suggests that, if sequencing had been deeper, complete capture of the 180 taxa in the target mock community might have been possible. Together, these results suggest that the choice of primer can be paramount when focusing on the LSU region.

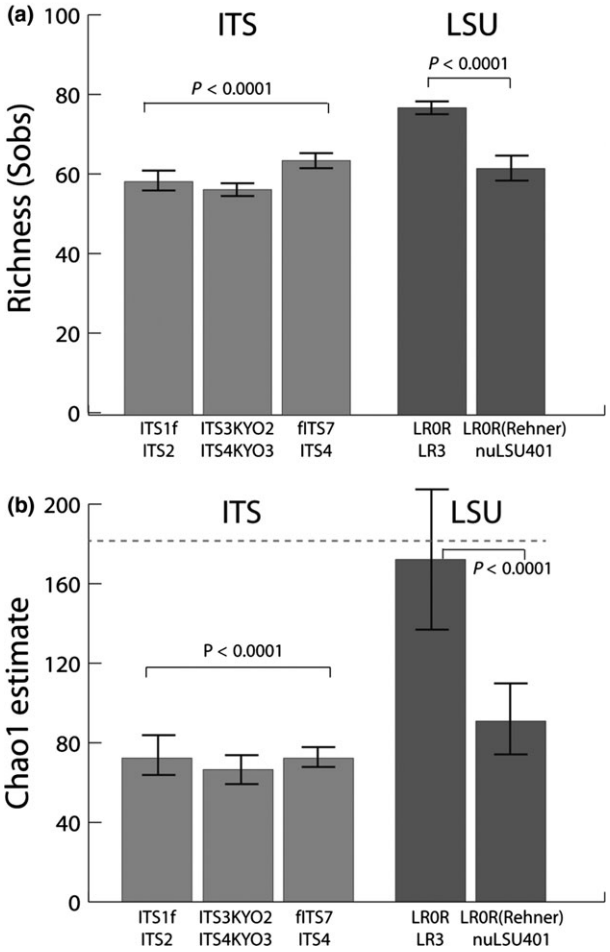

Fig. 2 Estimates of (a) Observed Richness (S_{obs}) and (b) minimum richness estimates (Chao1) means of the mock community for ITS (left) and LSU (right) primer pairs after subsampling 500 sequences for 1000 iterations (error bars are standard deviation). For both (a) and (b), all primer pair estimates were significant. Horizontal dashed line in (b) is the number of unique taxa in the mock community.

Discussion

Experiments on microbial community ecology through sequencing are often conducted using common 'universal' primer pairs from the literature. We suggest that it can be fruitful to test potential primer pairs against mock communities comprised of isolated culture representatives, to help ensure the suitability of a primer pair prior to launching an expensive full-scale community sequencing effort. Such an approach may help ensure that the communities represented by sequencing capture the most diversity present in the target communities, whilst minimizing potential skewness in taxonomic distribution.

Here, we demonstrate a cost-effective, efficient and high-throughput method to optimize primer selection utilizing a locally isolated mock community to inform downstream sequence generation and analyses (Fig. 3).

It is noteworthy that whilst we utilized a detailed culture library for our mock community, this is not always required. Our isolates are voucher quality because phylogenetic placement is ongoing; however, one could

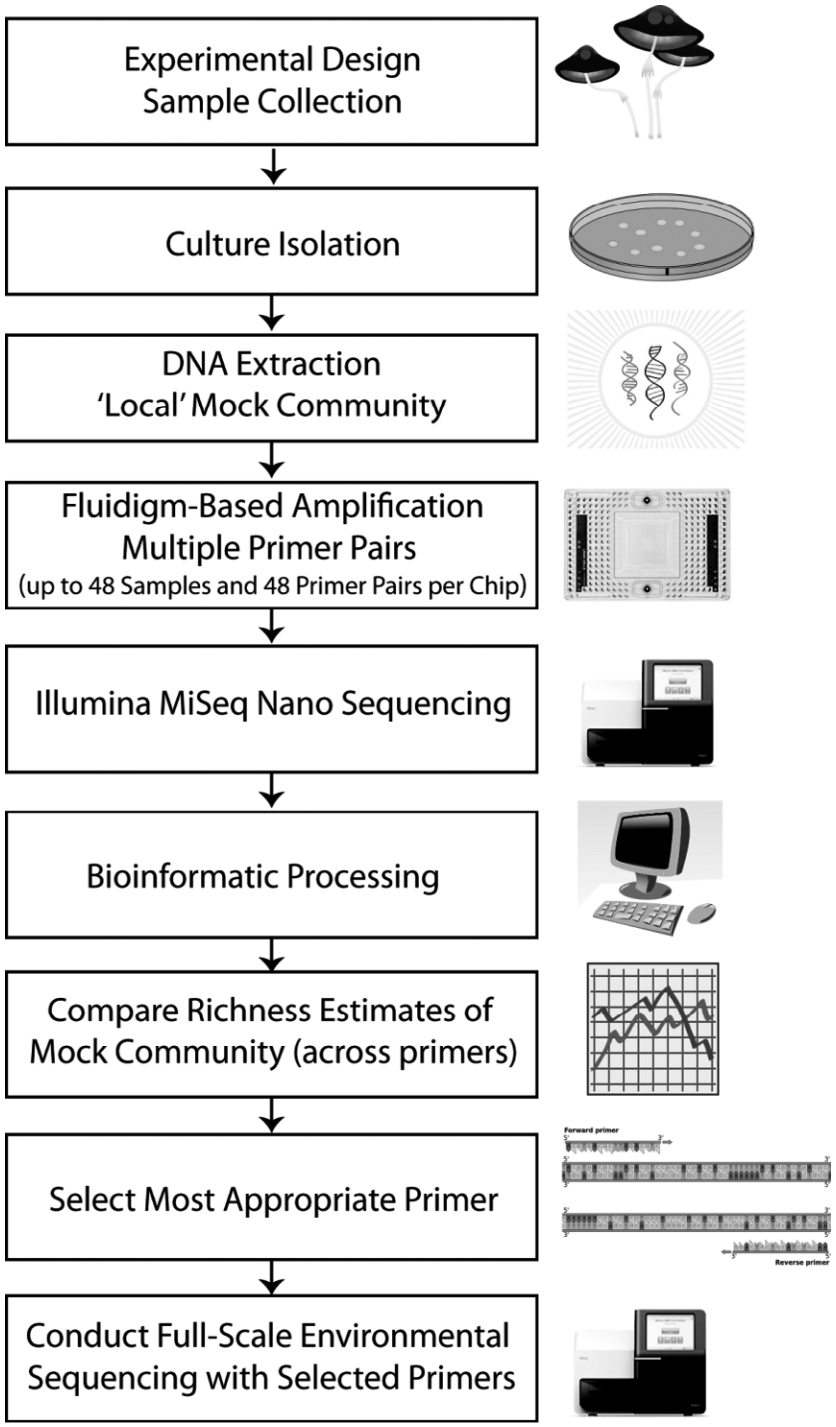

Fig. 3 Workflow depicting the steps involved in using Fluidigm IFC chips to test primer efficiencies. All images in this figure are in the public domain under Creative Commons CC0 1.0. Images of the Fluidigm IFC chip and Illumina MiSeq sequencer are taken directly from their respective websites.

isolate samples for mock community analyses using a less detailed approach such as mixed isolation and streaking for single-cell isolates. Additionally, in the absence of experimentally derived isolates, an *in silico* approach could be adopted to aid in primer selection (see Angly *et al*. 2012; Bokulich & Mills 2013), but such an approach may lack the system specificity afforded by the simultaneous isolation of cultures alongside community genomic DNA. Also, the optimal primer pair might vary among systems and questions; therefore, it becomes crucial to make informed primer decisions on an experiment-by-experiment basis. For our study system, fungi associated with wood submerged in Panamanian rivers, the results here from both observed and extrapolated minimum richness suggest that, if the ITS regions were the preferred target, then the primers

ITS1f-ITS2 and fITS7-ITS4 would both be suitable because they captured similarly unbiased taxonomic distributions and similar amounts of diversity from the culture-based local mock community. Similarly, if targets on the LSU are preferred, the primers LR0R-LR3 outperformed the other LSU primers (and the ITS primers), capturing the most diversity with minimal organismal biases. It is important to note that there are myriad LSU or ITS primers beyond the most commonly used sets that were tested here and that some of these untested primers might be even more appropriate for our system.

With the advent of next-generation sequencing and its resulting revolution of ecology come novel challenges including computational limitations, optimization of bioinformatics methods and incorporation of complex environmental measurements into community analyses. These rapidly advancing molecular methods have drastically increased our understanding of community ecology, and studies on microbial ecology and the utilization of molecular methodologies will likely remain coupled into the future (Peay *et al.* 2008). Many aspects of experimental design can ensure the most unbiased and accurate analysis possible of the queried community, and there has been much focus to date on optimizing how the raw sequence data are screened and clustered to include only the cleanest sequences for drawing ecological conclusions (*e.g.* Huse *et al.* 2010; Kunin *et al.* 2010; Tedersoo *et al.* 2010; Bokulich *et al.* 2013; Brown *et al.* 2015a,b). These post-sequencing technical investigations are crucial to environmental sequencing efforts, but equally important are presequencing considerations such as proper experimental design (Lindahl *et al.* 2013) and amplicon generation methods (*see* Bazzicalupo *et al.* 2013; Blaalid *et al.* 2013; Oliver *et al.* 2015).

Our results suggest that simple decisions, such as selection of primers used for amplification, may have large impacts on the study outcome. Though potentially influential, the informed choice of primers is often neglected (Bokulich & Mills 2013). Primers can skew the reported taxonomic distributions; nevertheless, general ecological patterns using different universal primers are generally consistent even though the magnitude and directionality of these shifts may differ (Blaalid *et al.* 2013; Brown *et al.* 2014). These effects might also depend on the taxonomic scale at which the data are examined. Our data would suggest that, even when broad scale detection of community-wide shifts are likely to remain unchanged, finer-scale patterns – that is, shifts of abundance at the OTU, genus, or family level – are likely to be strongly affected by primer bias.

In our study, the LSU primer pair LR0R-LR3 provided the best coverage of the mock community in terms of highest observed richness, extrapolated minimum richness estimates and recovery of the most taxonomic breadth in the collector's curves. The use of richness estimators to determine suitability of primer selection is not without issue, namely richness estimates do not account for hetero/homogeneity of the taxa distributions (*i.e.* sample beta diversity and evenness). Additionally, one cannot be certain how such a method as described here scales to a full community analysis as one cannot predict the capture of unculturable taxa not encompassed within the mock community. Nevertheless, the ability to capture a larger proportion of known mock community members is a simple and robust way to inform primer choice prior to a large-scale sequencing investment. Whilst the criteria for choosing primers and even the gene region to target (*i.e.* ITS or LSU) will remain particular to the study's experimental goals, utilizing the relatively inexpensive combination of Fluidigm Access Arrays with a singular Illumina HiSeq Nano reaction can allow researchers to *a priori* optimize quality data generation and thus downstream ecological inferences. We hope that this strategy can help researchers avoid placing of all their eggs in one basket (one primer set), which may result in inefficient sequence yields and/or skewed representation of reported taxonomic distributions from environmental data.

Conclusions

Here, we demonstrate the use of Fluidigm to optimize primer selection for microbial community analyses by utilizing a taxonomically similarly distributed mock community isolated from the same submerged substrate. Using pure cultures from the same sampling regime as the environmental DNA sampling allows for a unique ability to ensure that we are capturing environmental taxa most efficiently, as the mock community isolated from the same field collection will likely share much of the same inherent organismal primer biases. Although primer biases will likely never be eliminated, it is crucial to optimize sequence-based community ecology methods to minimize taxonomic biases to allow for informed inferences and interpretations of observed community ecological patterns.

Acknowledgements

We are greatly appreciative to everyone associated with the implementation and collection of these samples in Panama, particularly Mabelle Chong, Sergio Mosquera and Yaravi Suarez. We thank Mark Band and Chris Wright of the Roy J. Carver Biotechnology Center at the University of Illinois for technical assistance and sequence production. This work was funded through a Dimensions of Biodiversity Grant from the National Science Foundation (DEB 1241212 to Ferrer, Dalling and Heath).

References

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does sequence abundance count? *Molecular Ecology*, **19**, 5555–5565.

Anderson IC, Cairney JWG (2004) Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. *Environmental Microbiology*, **6**, 769–779.

Angly FE, Willner D, Rohwer F, Hugenholtz P, Tyson GW (2012) Grinder: a versatile amplicon and shotgun sequence simulator. *Nucleic Acid Research*, **40**, e94. doi: 10.1093/nar/gks251.

Bazzicalupo AL, Bálint M, Schmitt I (2013) Comparison of the ITS1 and ITS2 rDNA in 454 sequencing of hyperdiverse fungal communities. *Fungal Ecology*, **6**, 102–109.

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

Blaalid R, Kuman S, Nilsson RH, Abarenkov K, Kirk PM, Kauderud H (2013) ITS1 versus ITS2 as DNA metabarcodes for fungi. *Molecular Ecology Resources*, **13**, 218–224.

Bokulich NA, Mills DA (2013) Improved selection of Internal Transcribed Spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Applied and Environmental Microbiology*, **79**, 2519–2526.

Bokulich NA, Sabramanian A, Faith JJ *et al.* (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, **10**, 57–59.

Brown SP, Jumpponen A (2015) Phylogenetic diversity analyses reveal disparity between fungal and bacterial communities during microbial primary succession. *Soil Biology & Biochemistry*, **89**, 52–60.

Brown SP, Callaham MA, Oliver AK, Jumpponen A (2013) Deep Ion Torrent sequencing identifies soil fungal community shifts after frequent fires in a southeastern US forest ecosystem. *FEMS Microbiology Ecology*, **86**, 557–556.

Brown SP, Ridgon-Huss AR, Jumpponen A (2014) Analyses of ITS and LSU gene regions provide congruent results on fungal community responses. *Fungal Ecology*, **9**, 65–68.

Brown SP, Olson BJSC, Jumpponen A (2015a) Fungi and algae co-occur in snow: an issue of shared habitat or algal facilitation of heterotrophs? *Arctic, Antarctic, and Alpine Research*, **47**, 727–747.

Brown SP, Veach AM, Rigdon-Huss AR *et al.* (2015b) Scraping the bottom of the barrel: are rare high throughput sequences artifacts? *Fungal Ecology*, **13**, 221–225.

Caporaso JG, Lauber CL, Walters WA *et al.* (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, USA*, **108**, 4516–4522.

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

Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783–791.

Colwell RK (2006) EstimateS: Statistical estimation of species richness and shared species from samples. Version 9 and earlier. User's guide and application. Available from http://http://viceroy.eeb.uconn.edu/estimates/

Dörning H, Clerc P, Grube M, Wedin M (2000) Mycobiont-specific PCR primers for the amplification of nuclear ITS and LSU rDNA from lichenized ascomycetes. *Lichenologist*, **32**, 200–204.

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

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

Hugenholtz P, Pace NR (1996) Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends in Biotechnology*, **14**, 190–197.

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

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

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

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

Jumpponen A, Brown SP, Trappe JM, Cázares E, Strömmer R (2015) Analyses of sporocarps, morphotyped ectomycorrhizae, environmental ITS and LSU sequences identify common genera that occur at a periglacial site. *Journal of Fungi*, **1**, 76–93.

Köljalg U, Nilsson RH, Abarenkov K *et al.* (2013) Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, **22**, 5271–5277.

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

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

Lothamer K, Brown SP, Mattox JD, Jumpponen A (2014) Comparison of root-associated communities of native and non-native ectomycorrhizal hosts in an urban landscape. *Mycorrhiza*, **24**, 267–280.

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

Mueller RC, Gallegos-Graves LV, Kuske CR (2016) A new fungal large subunit ribosomal RNA primer for high throughput sequencing surveys. *FEMS Microbiology Ecology*. doi:10.1093/femsec/fiv153.

Needleman SB, Wunsch DC (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, **48**, 443–453.

Nguyen NH, Smith D, Peay K, Kennedy P (2015) Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytologist*, **205**, 1389–1393.

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

Nilsson RK, Tedersoo L, Abarenkov K *et al.* (2012) Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. *MycoKeys*, **4**, 37–63.

Oliver AK, Brown SP, Callaham MA, Jumpponen A (2015) Polymerase matters: non-proofreading enzymes inflate fungal community richness estimates by up to 15%. *Fungal Ecology*, **15**, 86–89.

Peay KG, Kennedy PG, Bruns TD (2008) Fungal community ecology: a hybrid beast with a molecular master. *BioScience*, **58**, 799–810.

Porras-Alfaro A, Liu KL, Kuske CR, Xie G (2014) From genus to phylum: large-subunit and internal transcribed spacer rRNA operon regions show similar classification accuracies influenced by database composition. *Applied and Environmental Microbiology*, **18**, 829–840.

Rehner SA, Samuels GJ (1994) Taxonomy and phylogeny of *Gliocladium* analyzed from nuclear large subunit ribosomal DNA sequences. *Mycological Research*, **98**, 625–634.

Renker C, Heinrichs J, Kaldorf M, Buscot F (2003) Combining nested PCR and restriction digest of the internal transcribed spacer region to characterize arbuscular mycorrhizal fungi on roots from the field. *Mycorrhiza*, **13**, 191–198.

Schloss PD (2013) Secondary structure improves OTU assignments of 16S rRNA gene sequences. *The ISME Journal*, **7**, 457–460.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mother: open-source, platform-independent, community supported software for

describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.
Schmidt TSB, Rodrigues JFM, von Mering C (2015) Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environmental Microbiology*, **17**, 1689–1706.
Schoch CL, Seifert KA (2012) Reply to Kiss; Internal transcribed spacer (ITS) remains the best candidate as a universal DNA barcode marker for *Fungi* despite imperfections. *Proceeding of the National Academy of Science, USA*, **109**, E1812.
Schoch CL, Seifert KA, Huhndorf S *et al*. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal NA barcode marker for *Fungi*. *Proceedings of the National Academy of Science, USA*, **109**, 6241–6246.
Seifert KA (2009) Progress toward DNA barcoding of fungi. *Molecular Ecology Resources*, **9**(s1), 83–89.
Singh P, Raghukumar C, Meena RM, Verma P, Shouche Y (2012) Fungal diversity in deep-seal sediments revealed by culture dependent and culture-independent approaches. *Fungal Ecology*, **5**, 543–553.
Sogin ML, Morrison HG, Huber JA *et al*. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences, USA*, **103**, 12115–12120.
Tedersoo L, Nilsson RH, Abarenkov K *et al*. (2010) 454 pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist*, **188**, 291–301.
Toju H, Tanabe AS, Yamamoto S, Sato H (2012) High-coverage ITS primers for the DNA-based identification of Ascomycetes and Basidiomycetes in environmental samples. *PLoS ONE*, **7**, e40863.
Vaz-Moreira I, Egas C, Nunes OC, Manaia CM (2011) Culture-dependent and culture-independent diversity surveys target different bacteria: a case study in a freshwater sample. *Antonie van Leeuwenhoek*, **100**, 245–257.
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.
Weber CF, Vilgalys R, Kuske CR (2013) Changes in fungal community composition in responses to elevated atmospheric CO_2 and nitrogen fertilization varies with soil horizon. *Frontiers in Microbiology*, **4**, 78.
Weiss S, Amir A, Hyde ER, Metcalf JL, Jin Song S, Knight R (2014) Tracking down the sources of experimental contamination in microbiome studies. *Genome Biology*, **15**, 564.
White TJ, Bruns T, Lee SJWT, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: A Guide to Methods and Applications* (eds Innis MA, Gelfand DH, Sninsky JJ & White TJ), pp. 315–322. Academic Press, Inc., New York, New York.
Xu Y, Moser C, Al-Soud WA *et al*. (2012) Culture-dependent and –independent investigation of microbial diversity on urinary catheters. *Journal of Clinical Microbiology*, **50**, 3901–3908.

A.F, J.W.D. and K.D.H. designed the experiment. A.F. and J.W.D. conducted field collections. S.P.B. and A.F. constructed and sequenced the mock community and S.P.B conducted all analyses. S.P.B and K.D.H wrote the manuscript, and all authors edited the manuscript.

Data accessibility

Paired fastq sequences of the mock community are accessioned under Sequences Read Archive (SRA) at NCBI (BioProject PRJNA307545, BioSample SAMN04282889, Experiment SRX1515368, Run SRR3083895. Sanger sequences of the mock community are deposited in GenBank at NCBI (Accession nos KU535697-KU535859; Dryad DOI:10.5061/dryad.tn6vn). Representative sequences of OTU for each Primer pair are accessioned on Dryad Digital Repository (DOI:10.5061/dryad.tn6vn)

Supporting Information

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

Table S1 Best BLASTn matches of 166 single spore isolates of the mock community against GenBank limited to only typified material.

Table S2 BLASTn scores of each representative OTU sequences for each tested primer pair, including closed matched mock community member (see Table S1).

Table S3 Iterative richness estimates (1000 iterations at a depth of 500 sequences per iteration) for OTU richness (S_{obs}) and Minimum extrapolated richness (Chao1) along with Standard Deviations (SD) of these estimates.