Novel ITS1 Fungal Primers for Characterization of the Mycobiome

Mykhaylo Usyk,[a] Christine P. Zolnik,[a,b] Hitesh Patel,[c] Michael H. Levi,[c,d]
Robert D. Burk[a,e]

Department of Pediatrics, Albert Einstein College of Medicine, Bronx, New York, USA[a]; Department of Biology, Long Island University, Brooklyn, New York, USA[b]; Department of Pathology, Montefiore Medical Center, Bronx, New York, USA[c]; Department of Pathology, Albert Einstein College of Medicine, Bronx, New York, USA[d]; Departments of Obstetrics & Gynecology and Women's Health, Epidemiology and Population Health, and Microbiology & Immunology, Albert Einstein College of Medicine, Bronx, New York, USA[e]

ABSTRACT Studies of the human microbiome frequently omit characterization of fungal communities (the mycobiome), which limits our ability to investigate how fungal communities influence human health. The internal transcribed spacer 1 (ITS1) region of the eukaryotic ribosomal cluster has features allowing for wide taxonomic coverage and has been recognized as a suitable barcode region for species-level identification of fungal organisms. We developed custom ITS1 primer sets using iterative alignment refinement. Primer performance was evaluated using *in silico* testing and experimental testing of fungal cultures and human samples. Using an expanded novel reference database, SIS (18S-ITS1-5.8S), the newly designed primers showed an average *in silico* taxonomic coverage of $79.9\% \pm 7.1\%$ compared to a coverage of $44.6\% \pm 13.2\%$ using previously published primers ($P = 0.05$). The newly described primer sets recovered an average of $21,830 \pm 225$ fungal reads from fungal isolate culture samples, whereas the previously published primers had an average of $3,305 \pm 1,621$ reads ($P = 0.03$). Of note was an increase in the taxonomic coverage of the *Candida* genus, which went from a mean coverage of $59.5\% \pm 13\%$ to $100.0\% \pm 0.0\%$ ($P = 0.0015$) comparing the previously described primers to the new primers, respectively. The newly developed ITS1 primer sets significantly improve general taxonomic coverage of fungal communities infecting humans and increased read depth by an order of magnitude over the best-performing published primer set tested. The overall best-performing primer pair in terms of taxonomic coverage and read recovery, ITS1-30F/ITS1-217R, will aid in advancing research in the area of the human mycobiome.

IMPORTANCE The mycobiome constitutes all the fungal organisms within an environment or biological niche. The fungi are eukaryotes, are extremely heterogeneous, and include yeasts and molds that colonize humans as part of the microbiome. In addition, fungi can also infect humans and cause disease. Characterization of the bacterial component of the microbiome was revolutionized by 16S rRNA gene fragment amplification, next-generation sequencing technologies, and bioinformatics pipelines. Characterization of the mycobiome has often not been included in microbiome studies because of limitations in amplification systems. This report revisited the selection of PCR primers that amplify the fungal ITS1 region. We have identified primers with superior identification of fungi present in the database. We have compared the new primer sets against those previously used in the literature and show a significant improvement in read count and taxon identification. These primers should facilitate the study of fungi in human physiology and disease states.

KEYWORDS ITS1, yeast, fungi, mycobiome, oral, primer design

As innovations in the field of next-generation sequencing (NGS) progress and high-throughput bioinformatic analyses become more prevalent, the microbiome has emerged as a field of increasing importance. Microbiome studies have primarily focused on the identification of significant bacterial taxa or community states in human- and animal-pathogenic conditions because of the ease of community characterization by sequencing PCR-amplified fragments of the prokaryotic 16S rRNA gene (1). This gene is ideal for bacterial identification due to its universal presence in pro-karyotes, homologous structure, and evolutionary relatedness allowing taxonomic reconstruction. Although fungi have been shown to be important in human health due in part to their ability to secrete highly active metabolites that are directly involved in pathogenesis (2) and indirectly through modulation of the microbiome (3), the study of pathogenic and nonpathogenic fungi and fungal communities in general has lagged behind studies of bacteria (4, 5).

Unlike bacteria, fungal evolution and morphological diversity are not intimately associated with single genetic markers. The morphological system that has primarily been used to classify fungi presents a barrier for researchers by including multiple potential species names for the same organism. For example, morphological differences present throughout the various reproductive stages of fungi (i.e., anamorph and telemorph) (6) may result in multiple taxonomic classifications for the same organism. In contrast, molecular methods that target specific genetic regions such as cytochrome *c* oxidase subunit 1 (CO1) (7) and the internal transcribed spacers (ITSs) (see Fig. 1) of the eukaryotic ribosomal cluster (8), provide a new paradigm for fungal taxonomy. Placement of the same organism into multiple groups due to morphological characterization confounds molecular classification, especially when there is significant genetic diversity in the organism of interest. Investigators are therefore limited to the use of genetic markers that are informative for species-level identification, but unfortunately lack the ability to perform phylogenetic reconstruction given the complex evolutionary history of fungi (4). A currently proposed barcode DNA region for fungal community identification is the ITS1 region, which allows for species-level resolution in a large number of fungi and provides amplicon sizes suitable for current short read NGS platforms like Illumina (9).

Despite the availability of a suitable DNA region for fungal community analyses, studies of medically relevant fungi use previously developed, "universal" ITS1 primers that are limited by taxonomic bias as they were developed based on small subset of soil fungi available at the time of their creation and commonly generate a low number of sequence reads (10–14). The latter is especially important because community-based analyses suffer when a sample does not achieve a minimum read threshold necessary to characterize the community—a threshold readily identifiable through rarefaction analysis (15). The present study utilizes bioinformatics and an iterative approach to identify fungal sequences for primer design for broad taxonomic coverage using the ITS1 barcode region. This study further expands on the currently available tools by creating a database of contiguous 18S-ITS1-5.8S (SIS) sequences, which are omitted from assembled fungal reference genomes and makes it possible to bioinformatically evaluate primer pairs. In addition to *in silico* evaluation against the SIS database, the primers were empirically tested with fungal cultures and clinical samples. We provide data that these primers significantly improve fungal taxonomic coverage, fungal read recovery, and accuracy of characterization of human-associated fungal communities.

RESULTS

ITS1 primer design. A total of seven (five forward and two reverse) published primers amplifying the ITS1 region were initially chosen for further analysis based on reported overall taxonomic coverage and amplicon size (see Table S1 in the supplemental material). Additionally, 85 primers (78 forward and 7 reverse) were custom designed using the SIS database (see Table S2 in the supplemental material). Three forward primers and one reverse primer representing the previously published primers and four forward primers and one reverse primer representing the custom primers were
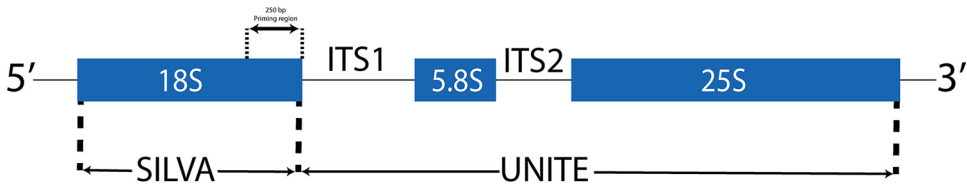
FIG 1 ITS rRNA gene locus. Schematic of the eukaryotic ribosomal gene cluster. The SILVA database contains sequences of the 18S gene, while the UNITE database contains sequences from the ITS1-5.8S-ITS2-25S rRNA gene cluster (not to scale). For development of our custom primers, we created a merged SILVA and UNITE database to simulate the 18S-ITS1-5.8S region. A 250-bp region at the 3′ end of the 18S gene was individually isolated when designing the forward primers.

further evaluated by extended *in silico* testing using the SIS database and experimental testing of fungal cultures (Table 1). The four custom-designed forward primer pairs were also tested with one published reverse primer (Table 1). The chosen primer pairs were selected based on high taxonomic coverage of sequences in the UNITE (User-friendly Nordic ITS Ectomycorrhiza) database, in particular for medically relevant cervicovaginal taxa (not shown), as well as a <5°C difference in melting temperature (T_m) between the forward and reverse primers.

In silico analyses. *In silico* testing demonstrated that the custom-designed primers achieved an average *in silico* taxonomic coverage of 79.9% ± 7.1%, while the published primer pairs had an average coverage of 44.6% ± 13.2% ($P = 0.05$) (Fig. 2 and 3). Both reverse primers tested demonstrated similar performance, covering nearly all reference fungi, with coverage of 90.0% ± 14.8% and 87.1% ± 18.9% ($P = 0.39$) for the custom and published primers, respectively. This is likely due to the high conservation of the flanking 5.8S region (Fig. 1) compared to the larger and more variable 18S target of the forward primers. Coverage across the analyzed fungal phyla and *Candida* is summarized in Fig. S2 in the supplemental material. The custom primers were able to improve coverage of the medically relevant fungal genus *Candida* by being able to produce amplicons for all of its species contained in the SIS database compared to an average coverage of 59.5% ± 13% from the previously reported primers.

Testing of fungal cultures. The 13 fungal culture isolates used in this study represent common taxa found in the cervicovaginal tract and include species from two phyla (*Ascomycota* and *Basidiomycota*), six classes, and seven orders (see Table S3 in the supplemental material). Read recovery analysis for the selected primer pairs demonstrated that the custom-designed primer sets resulted in an order of magnitude increase over all but one published primer pair from the literature [designated by "(L)"]: there was a 3-fold increase over the ITS5 (L) primer (Fig. 4). The newly designed primers were able to achieve a mean sequencing depth of 21,830 ± 225 fungal reads, while the published primers amplified an average of 3,305 ± 1,621 fungal reads ($P = 0.029$). In

TABLE 1 Published and custom-designed primers used for *in silico* and experimental testing

Primer	Direction	Primer sequence (5′ to 3′)	Reference(s)
Published			
ITS1 (L)	Forward	TCCGTAGGTGAACCTGCGG	11
ITS1-F (L)	Forward	CTTGGTCATTTAGAGGAAGTAA	20–23
ITS5 (L)	Forward	GGAAGTAAAAGTCGTAACAAGG	11
ITS2 (L)[a]	Reverse	GCTGCGTTCTTCATCGATGC	11, 21, 23, 28–35
Custom designed			
ITS1-27F (N)	Forward	TACGTCCCTGCCCTTTGTAC	
ITS1-30F (N)	Forward	GTCCCTGCCCTTTGTACACA	
ITS1-34F (N)	Forward	CTGCCCTTTGTACACACCGC	
ITS1-48F (N)	Forward	ACACACCGCCCGTCGCTACT	
ITS1-217R (N)	Reverse	TTTCGCTGCGTTCTTCATCG	

[a]This published reverse primer was also tested against the four selected custom-designed forward primers.
"(L)" designates primers from the literature, and "(N)" indicates primers newly designed for this study.
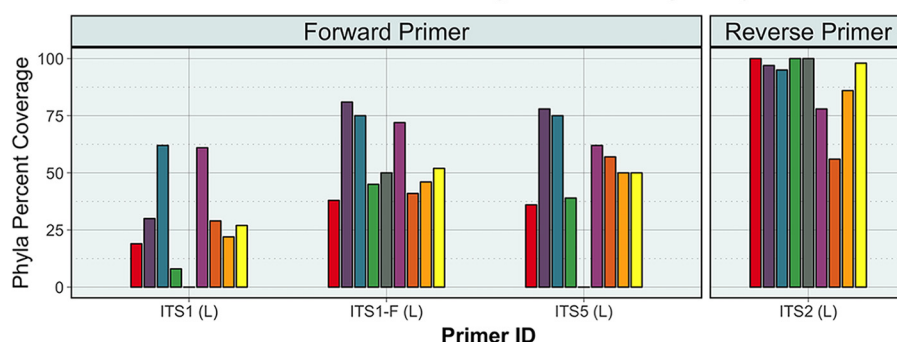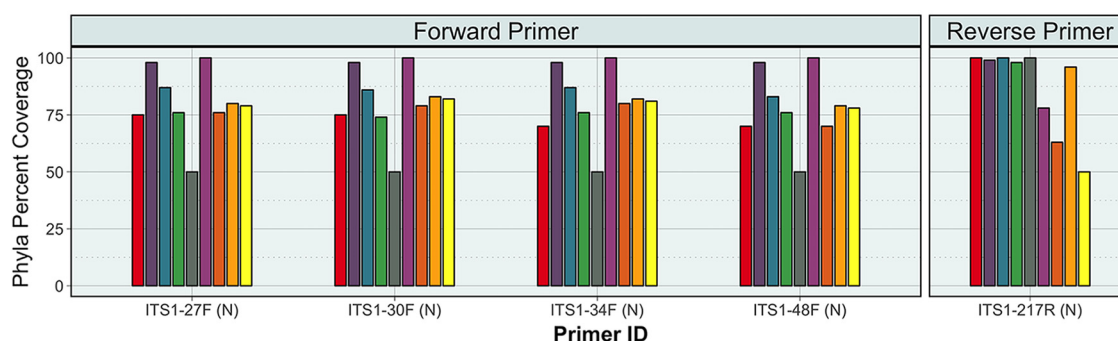
A. *In silico* Literature Primer Coverage Across Fungal Phyla

B. *In silico* Custom Primer Coverage Across Fungal Phyla

FIG 2 *In silico* coverage across fungal phyla. Predicted taxonomic coverage was assessed using PrimerProspector. (A) The forward literature fungal primers ITS1 (L), ITS1-F (L), and ITS5 (L) had significantly lower overall taxonomic coverage than the (B) newly created forward primers ITS1-27F (N), ITS1-30F (N), ITS1-34F (N), and ITS1-48F (N). The custom-designed reverse primer ITS1-217R (N) and the published reverse primer ITS2 (L) both demonstrated high predicted taxonomic coverage across the phyla.

addition to an increase in sequencing depth, there was a more consistent read recovery between the newly designed primer set [designated by "(N)"] than with the literature set. The 20 PCR-negative controls averaged 7 ± 2 fungal reads.

Across the 13 fungal cultures (Table 2), the custom-designed primers provided an overall improvement approaching statistical significance ($P = 0.057$) in correct identification compared to the published primers. A culture was considered to be correctly identified if at least 50% of the reads from a given primer pair matched the species identification of the culture, the anamorph of the cultured species, or higher taxonomic strata that the cultured species belongs to. Custom primers were able to identify all cultures. The worst-performing primer was the published forward primer ITS1-F (L), which resulted in identification of 10 out of 13 cultures, while primers ITS1 (L) and ITS5 (L) both correctly identified 11 of the 13 cultures. This misidentification occurred for cultures when the primers had less than 1,000 reads, highlighting the importance of adequate sample sequencing depth.

Clinical sample analyses. Reads from clinical samples were bioinformatically processed to remove any reads that were not identified as being fungal. Following this filtering, the clinical samples averaged 10,068 ± 1,227 fungal reads per sample. The read breakdown for each body site was as followed: the cervical samples had 6,221 ± 638 reads, whereas samples from the oral and anal sites had similar coverage of 11,357 ± 1,677 and 12,626 ± 3,174 fungal reads, respectively.

Alpha rarefaction using the Shannon metric was used to determine whether the fungal community was adequately represented for each sample (Fig. 5). The majority of the sample-primer combinations achieved adequate sequencing depth across the three body sites: 74/84 achieved sufficient sampling depth with the plateau point being designated at 500 reads.
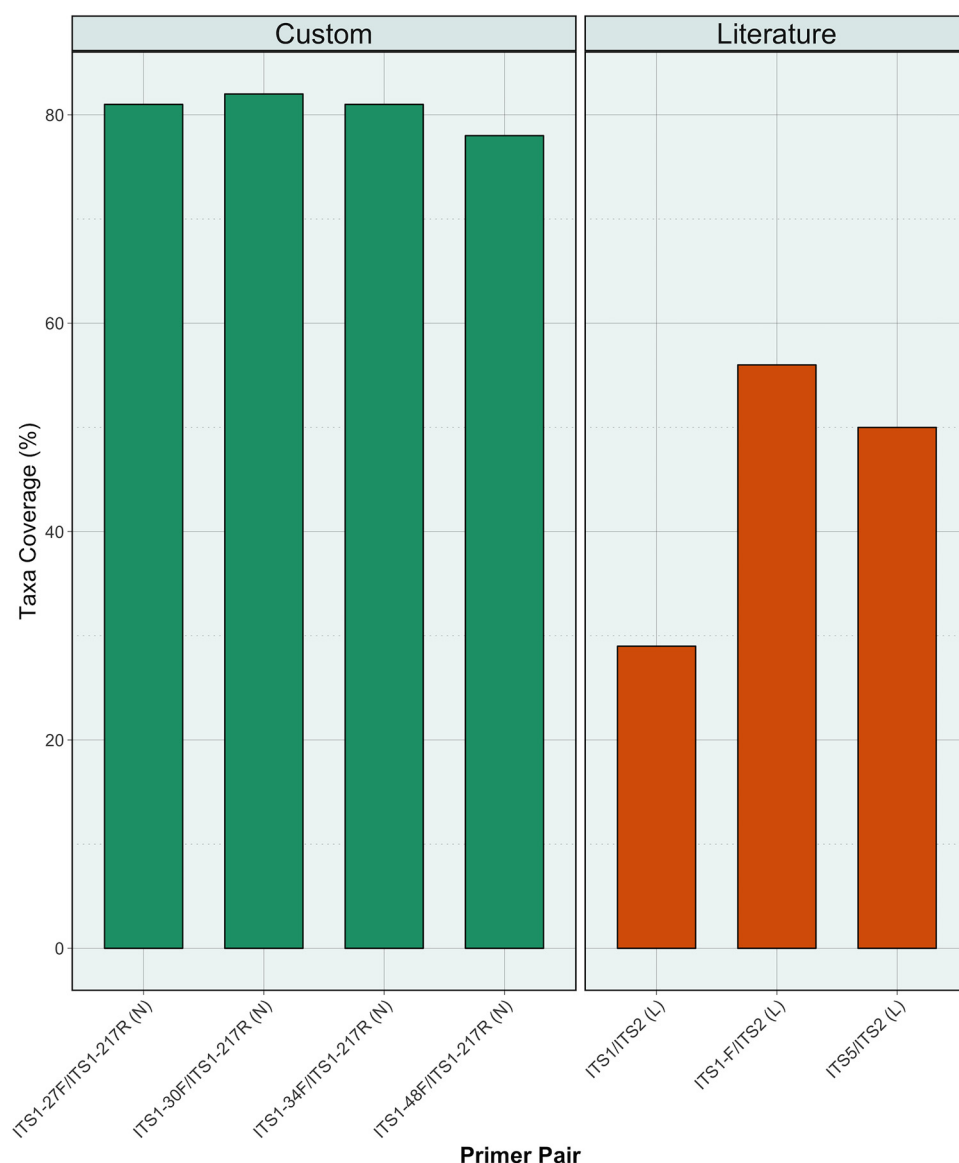
In silico Taxa Coverage - Primer Pairs

FIG 3 *In silico* taxon coverage. The average *in silico* taxonomic coverage estimates obtained with PrimerProspector for the selected forward primers from the literature [designated "(L)"] and newly designed forward primers [designated "(N)"]. The published ITS1 (L) primer was recommended by UNITE and is therefore shown in this figure despite the low coverage performance. Coverage estimates are based on 5,789 simulated species in the SIS database. Coverage estimates are based on the default alignment criteria of PrimerProspector.

DISCUSSION

Most mycobiome studies that utilize next-generation sequencing have typically used primer sets that were developed with a limited reference set of fungal species (11). Using *in silico* testing with a novel database (SIS), we determined that frequently cited primer sets rarely exceed 50% taxonomic coverage of fungal sequences in the UNITE database and result in low numbers of sequence reads during NGS runs (Fig. 3 and 4). As a result, these published primer sets limit our ability to adequately evaluate fungal communities. We determined using *in silico* testing and the newly designed SIS database that our newly designed primers are able to amplify 79.9% ± 7.1% of the species currently present in the UNITE database. In addition, we found using a sequencing experiment with fungal isolates that the redesigned primers increase fungal read recovery from 3,305 ± 1,621 reads to 21,830 ± 225 reads. We further show that
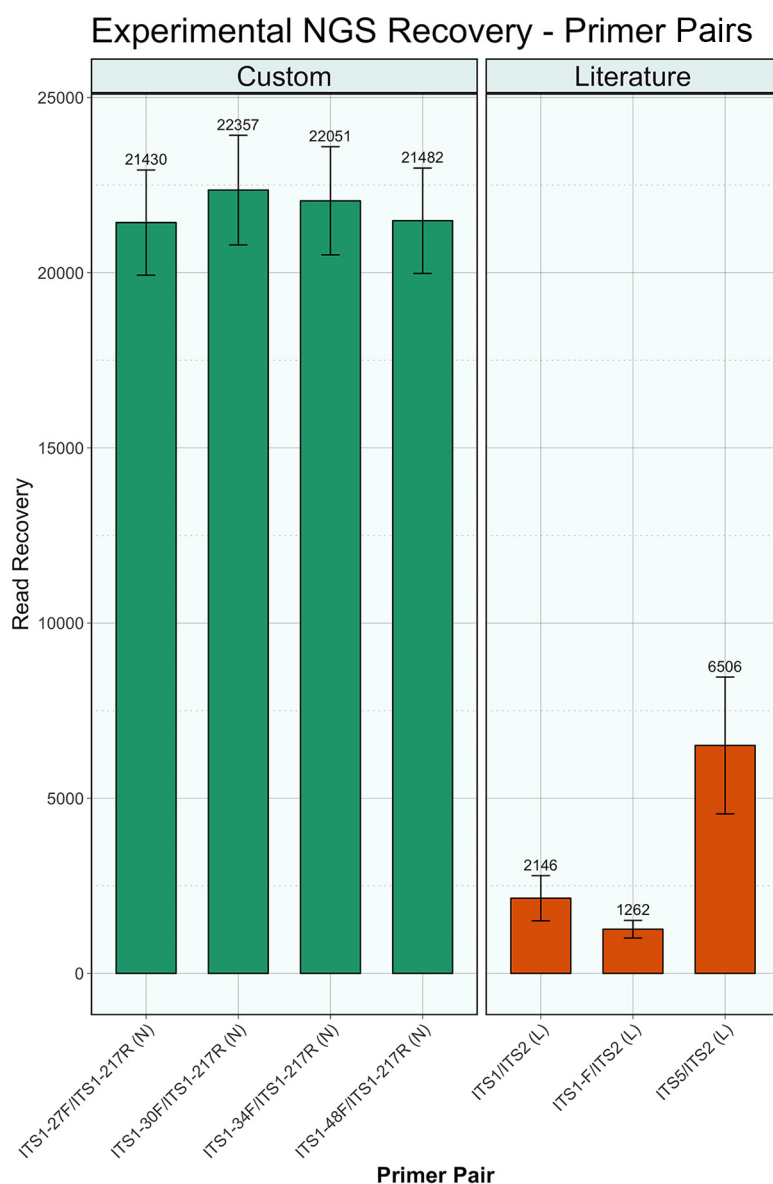
FIG 4 Experimental NGS read recovery. Average read recovery for the selected newly designed [designated "(N)"] and published forward primers from the literature [designated "(L)"] is shown at the top of each colored bar. Error bars show the standard error of the mean. The names of the primer pairs are indicated on the *x* axis.

the increased sequencing depth is such that it can capture the fungal diversity at three body sites. The only exception to this is the newly designed primer pair ITS1-34F/ITS1-217R (N), which failed for several samples in the anal region. The loss is likely due to technical or sequencing issues, as opposed to the limitation of the primers to amplify fungi from this anatomic region. A technical issue is more likely to be the case, rather than an inability of the custom primers to amplify anal fungi because nonmetric multidimensional scaling (NMDS) analysis of the fungal communities across the sites indicates that the anal region overlaps both the oral and cervical regions, meaning that the regions have common species of fungi present (see Fig. S1 in the supplemental material) that were detected with this primer pair.

The primers described in this report were designed using a custom database (SIS) that combined information within the UNITE (16) and SILVA (17) databases. The SIS database made it possible to perform *in silico* PCR testing because, to the best of our knowledge, there were no available databases that contain the entire 18S-ITS1-5.8S

TABLE 2 Fungal species tested in this study representing common human cervicovaginal-associated species

Laboratory ID	Species
Fungal_01	*Saccharomyces cerevisiae*
Fungal_02	*Rhodotorula mucilaginosa*
Fungal_15	*Candida dubliniensis*
Fungal_18	*Candida tropicalis*
Fungal_31	*Cryptococcus neoformans*
Fungal_33	*Aspergillus flavus*
Fungal_35	*Aspergillus niger*
Fungal_36	*Aspergillus terreus*
Fungal_37	*Bipolaris* sp.
Fungal_38	*Microsporum canis*
Fungal_39	*Scedosporium apiospermum*
Fungal_56	*Candida albicans*
Fungal_58	*Candida parapsilosis*

contig. The SIS database will be publically available to facilitate further public ITS amplification testing. Although we showed experimentally that the simulated database allows for the development of significantly better primers over the current primer sets, it should be noted that the availability of the true reference contigs would make taxonomic coverage more accurate. Nevertheless, use of *in silico* analyses may have unexpected limitations that can be elucidated on experimental testing.

The primers developed with the help of the SIS database significantly increased taxonomic coverage and increased NGS yield by an order of magnitude for fungal cultures. The increased recovery of fungal reads in the human samples was most notable in the cervicovaginal region, which is known for low microbial biomass (bacteria and fungi) compared to other human body sites (18). High read counts are crucial to ensure that fungal communities are accurately characterized. The best overall primer set for the study of the fungal mycobiome in the cervicovaginal region was the ITS1-30F/ITS1-217R (N) pair (Table 1). This primer pair was the highest performing in
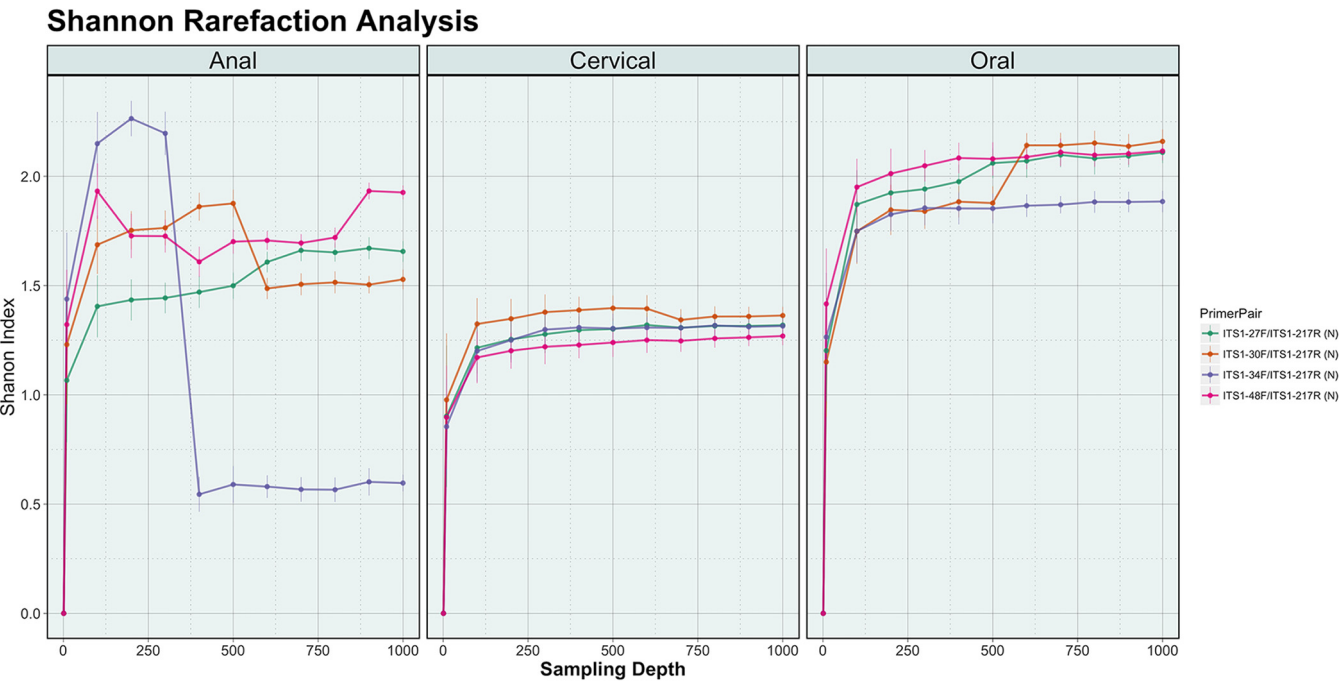
FIG 5 Shannon rarefaction analysis. Shown is Shannon alpha rarefaction analysis across three body sites for seven clinical samples using four newly designed primer pairs. Amplified samples were evaluated at depths of 1, 10, 100, 500, and 1,000 reads, with 100 replicates at each subsampling depth. Results for each primer pair with the samples were averaged and plotted by anatomic site. The primer pairs associated with each colored line are indicated in the key to the right of the figure.

terms of overall fungal culture identification and the best performing in terms of fungal read recovery of all tested primers (Fig. 4), which is critical when studying a low-mycobiome-biomass site. It should be reiterated, however, that the differences between the new primer sets were marginal, and any of the four custom-designed pairs offers a significant improvement in read recovery and taxonomic coverage. These newly designed primers will facilitate fungal microbiome (mycobiome) studies by expanding taxonomic coverage and ensuring the accuracy of community characterization by providing significant improvements in fungal sequencing depth. This is particularly an important improvement for studies of fungal communities in the cervicovaginal body region, where these types of studies are currently lacking (19).

This study was limited by the lack of an actual intact ITS region database—thus the need to design a custom database (SIS) of the 18S-ITS-5.8S contig to facilitate *in silico* evaluation of custom fungal primers. Although the developed primers seem to reflect an improvement over the literature primers in experimental analyses, having a true 18S-ITS-5.8S database would increase the accuracy of the results. The study may also be limited by the current set of fungal reference sequences. With constant improvements made to sequencing technology and constant updates made to fungal reference databases, it may be necessary to reexamine the newly designed primer sets to ensure that they can adequately capture the ever-expanding fungal diversity and if necessary create new primer-sets.

This study demonstrates significant limitation of commonly used ITS primers and presents newly designed primers that significantly expand the general taxonomic coverage and increase NGS read recovery. Additionally, the study presents a full 18S-ITS-5.8S contig database (SIS) that can be used to evaluate the priming efficacy of the ITS region and offers suggested parameters for running postsequencing operational taxonomic unit (OTU) clustering analysis.

MATERIALS AND METHODS

ITS1 primer assessment and custom primer design. Previously published ITS1 primers were identified from the literature [designated "(L)"], with an emphasis on studies of the human mycobiome and on fungal species of the terrestrial environment (Table S1) (11, 20–36). Additionally, primers recommended by the User-friendly Nordic ITS Ectomycorrhiza (UNITE) database were also considered (16). Primers from the literature and UNITE database were chosen for further investigation if they met the following criteria, based on published reports: (i) amplification of the ITS1 region, (ii) indication of utility as a general-purpose amplifier of the ITS1 locus without known large taxonomic biases, (iii) amplicon sizes below 700 bp and in which the predicted amplicon mean length for a given pair was 400 ± 100 bp, and (iv) evidence of coverage of known human-associated fungal taxa.

To develop new custom forward primers [designated "(N)"], a 250-bp region from the 3' end of the 18S fragment for each of the representative sequences in the SILVA high-quality rRNA database (17) was used to enable focus on the region immediately upstream of ITS1 (Fig. 1). These 18S subfragment sequences were aligned using MAFFT V7 (37). Of the 23,128 input sequences, some failed to align ($n = 63$) and were pruned from the input, and the entire set was realigned. The final aligned sequence set was then subjected to PrimerProspector (38) from the V1.9 release of QIIME virtualbox (39) to obtain degenerate forward primers (Table S2).

Reverse primers were created using the UNITE v7 database from the October 2014 release (10). The 5.8 S reference sequence fragment was selected in order to retrieve the contiguous ITS1 region with the flanking ribosomal sequences (Fig. 1). MAFFT v7 (37) was used for the alignment, and PrimerProspector was used to generate degenerate reverse primers. Due to the high conservation of the 5.8S region, no alignment pruning was needed.

18S-ITS1-5.8S (SIS) contig database creation. To test primers *in silico* and estimate taxonomic coverage, a contig of the 18S-ITS1-5.8S region was required (Fig. 1); however, a specific database that covers the entire region from 18S to 5.8S does not to our knowledge exist, nor is this region included in the published genome sequences of most fungi. We therefore designed an algorithm to merge sequences from SILVA and UNITE databases to simulate an intact 18S-ITS1-5.8S region. The custom database was designed to include the ITS1 locus with the entire 18S 5' region not just the targeted 250-bp 3' fragment, in order to ensure that the designed primers did not anneal to multiple conserved targets within the 18S gene. Taxonomic levels within each database are not in the same format (i.e., SILVA has more taxonomic levels than UNITE, which is fixed at seven levels); therefore, an iterative regular expression pattern-matching algorithm (40) was used to match the taxonomic descriptor for each UNITE sequence to a SILVA reference at the lowest possible taxonomic level. This procedure took advantage of the high conservation of the ITS1 flanking rRNA loci, which make these regions suitable to resolve fungi to the species level (41). The newly combined SILVA-UNITE database, termed SIS (18S̲-I̲TS1-5.8S̲), made it possible to then perform *in silico* PCR testing since the forward and reverse primer targets were present on each contig. The SIS database contains 5,789 simulated reference sequence contigs representing 3,842 fungal species.

In silico testing of primer sequences. Published and custom-designed primers (Tables S1 and S2) were evaluated to identify pairs with melting temperature differences of <5°C using OligoCalc (42). Next, the primer pairs were scored for taxonomic coverage and amplicon size using the SIS database with PrimerProspector (39). The amplicon range was chosen to specifically work with the MiSeq platform and ranged from 358 ± 83 bp to 384 ± 105 bp. Finally, the top-performing published primers from the literature (3 forward and 1 reverse) and newly designed primers (4 forward and 1 reverse) in terms of global fungal coverage across all phyla found in the UNITE v7 database (Table S3) were selected for experimental in silico testing (Table 1). Additionally, the final primer pairs (Table 1) were chosen to include taxonomic coverage of fungi that have been previously found in the cervicovaginal region.

Fungal cultures. Thirteen fungal culture isolates were obtained from two sources—the College of American Pathologists (CAP) samples (n = 2) and from the microbiology lab at the Montefiore Medical Center (n = 11) (Table 2). The Montefiore samples were subcultured on appropriate media for growth of fungus and the Montefiore Microbiology Laboratory classified isolated colonies. Yeast organisms were identified using the Phoenix automated system for identification, and rice Tween agar was used to evaluate yeast morphology and confirm identification. Filamentous fungi were identified using standard mycology techniques (i.e., microscopy and temperature requirements).

Clinical samples. Samples from seven anonymous individuals from our published studies were used (43, 44). Cervicovaginal and anal swab samples were collected in PreservCyt (Thin Prep, Marlborough, MA), and Scope mouthwash was used for oral sample collection. Samples were transferred to a 15-ml tube and gently centrifuged at 1,500 rpm for 5 min. After removing the supernatant by decanting, the pellet was rinsed in 3 ml of TE (10 mM Tris, 1.0 mM EDTA). This solution was then vortexed and centrifuged at 1,500 rpm for 5 min, and the supernatant was removed by decanting. The remaining pellet and any leftover solution (~150 μl) were stored at -20°C until further processing.

DNA extraction. DNA was extracted from all fungal cultures and clinical samples in a sterile biosafety cabinet. Fungal culture samples were incubated at 37°C for 30 min in an enzyme cocktail containing 15 μl of lysozyme (0.84 mg/ml; Sigma-Aldrich), 9 μl of mutanolysin (0.25 U/ml; Sigma-Aldrich), and 4 μl lysostaphin (21.10 U/ml; Sigma-Aldrich). The samples were then incubated at 56°C for 10 min following the addition of 15 μl proteinase K (20 mg/ml) and 150 μl buffer AL (Qiagen lysis buffer) and mixed by pulse-vortexing for 15 s. The samples were transferred to a screw-cap tube with 100 mg of UV-sterilized 0.1-mm-diameter zirconia-silica beads (11079101z; BioSpec, Bartlesville, OK) and beaten with a FastPrep-24 homogenizer (MP Biomedicals, Santa Ana, CA) at speed 6 for 40 s. Tubes were centrifuged at 750 × g for 30 s, 150 μl of supernatant was removed and placed in a new tube, and then 100 μl of 100% ethanol (EtOH) was added and mixed by pulse-vortexing for 15 s. After centrifugation at 750 × g for 30 s, the supernatant was added to the QIAamp mini-spin column (Qiagen, Valencia, CA) and centrifuged at 6,000 × g for 1 min. Column purification was performed according to the QIAamp DNA minikit directions starting at the AWI wash step, and DNA was collected in 100 μl of buffer AE (10 mM Tris, 0.5 mM EDTA [pH 9]).

DNA was extracted from the clinical samples using the QIAamp mini-spin column method (Qiagen) following the manufacturer's protocol. The purified DNA was eluted in 150 μl of elution buffer AE.

PCR and sequencing. The DNA from the fungal culture samples underwent PCR using three published primer pairs and eight primer pairs consisting of custom-designed plus published primers (Table 1). All primers were synthesized with Golay barcodes, providing unique dual barcodes for each PCR. For each primer pair, PCR protocols were optimized for annealing temperature using a Verti thermal cycler (Thermo Fisher, Waltham, MA). Based on these trials, PCR was performed in a 25-μl reaction mixture with 2.5 μl input of fungal culture DNA, 16.25 μl of double-distilled water (ddH$_2$O), 2.5 μl of USB 10× buffer with MgCl$_2$ (10 mM; Affymetrix, Santa Clara, CA), 1 μl of USB MgCl$_2$ (25 mM), 0.5 μl of deoxynucleoside triphosphate (dNTP) mixture (10 mM each; Roche Basel, Switzerland), 0.25 μl AmpliTaq Gold polymerase (5 U/μl; Applied Biosystems, Carlsbad, CA), 0.5 μl of Hotstart-IT DNA Fidelitaq polymerase (2.5 U/μl; Affymetrix), and 1 μl (5 μM) of each primer (IDT, Coralville, IA). Thermocycling was performed on a GeneAmp PCR system 9700 (Applied Biosystems) and included an initial denaturation of 95°C for 3 min, followed by 35 cycles of 95°C for 30 s, 55°C for 30 s, and 68°C for 2 min, followed by a final extension of 68°C for 10 min. Negative controls were 20 mock samples with all reagents, including barcoded primers but without any extracted DNA that went through PCR and NGS.

PCRs of the clinical samples were performed using four custom-designed primer pairs (Table 1). For these samples, PCRs were performed as described above using 10 μl of sample DNA.

PCR products for fungal cultures and clinical samples were pooled, and 100 μl of the pooled PCR products was loaded into a 4% agarose gel and run at 88 V for 4 h until the bands separated. The band for each primer pair was excised, purified with a QIAquick gel extraction kit (Qiagen), and quantified using a Qubit 2.0 fluorometric high-sensitivity double-stranded DNA (dsDNA) assay (Life Technologies, Inc., Carlsbad, CA).

Next-generation sequencing library preparation was performed using the KAPA LTP library preparation kit (KAPA Biosystems, Wilmington, MA) according to the manufacturer's protocol. The size integrity of the isolated amplicons was validated with a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) at the Genomics Core at Albert Einstein College of Medicine. High-throughput sequencing of libraries was carried out on an Illumina MiSeq (Illumina, San Diego, CA) with a 2× 300-bp paired-end read kit at the Genomics Core of the Albert Einstein College of Medicine.

Bioinformatic pipeline and statistical analyses. MiSeq sequence reads were demultiplexed based on the dual barcodes using Novocraft's Novobarcode V1.00 (45). Reads were trimmed for bases that fell below a PHRED score of 25 at the 3' end with PrinSeq V0.20.4 (46) and merged using PANDASEQ V1.20 (47) under the default settings.

Open reference OTU picking was employed using QIIME v1.9 open-reference OTU picking protocol and the UNITE database for the reference-based clustering component. VSEARCH v1.4.0 (48) was substituted for usearch within Mac Qiime to facilitate higher throughput, since VSEARCH allows all of a system's available memory to be used for processing. The OTU clustering threshold was set at 99% sequence identity to account for fungal heterogeneity as previously reported (16). Sequence dereplication and chimera removal were performed using the QIIME quality control protocol. Representative sequences for each OTU cluster were chosen based on sequence abundance. BLAST (49) was used to assign taxonomy using the UNITE database. The default behavior of BLAST in QIIME was changed to a minimum of 99% sequence identity for taxonomic assignment.

Data were processed in R version 3.3.1 (50). QIIME outputs were imported into R using the *phyloseq* (51) package and further processed with *vegan* (52), *coin* (53), and *reshape2* (54). Data visualization was performed using *ggplot* (55). A Wilcoxon-Mann-Whitney test, *stats* package, was used for assessment of the number of sequence read counts and taxonomic coverage for fungal cultures amplified with custom-designed and previously published primers.

Data availability. Raw data from the UNITE and SILVA reference databases are available for download from the source websites. The constructed SIS database is available for public access through the associated GitHub page: https://github.com/musyk07/18S-ITS1-5.8S-SIS-Database. The algorithm used to construct the SIS database is also available on GitHub.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSphere.00488-17.

FIG S1, TIF file, 2.2 MB.

FIG S2, TIF file, 2.6 MB.

TABLE S1, PDF file, 0.2 MB.

TABLE S2, PDF file, 0.1 MB.

TABLE S3, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A 108:4516–4522. https://doi.org/10.1073/pnas.1000080107.
2. Sexton AC, Howlett BJ. 2006. Parallels in fungal pathogenesis on plant and animal hosts. Eukaryot Cell 5:1941–1949. https://doi.org/10.1128/EC.00277-06.
3. Huffnagle GB, Noverr MC. 2013. The emerging world of the fungal microbiome. Trends Microbiol 21:334–341. https://doi.org/10.1016/j.tim.2013.04.002.
4. Seifert KA. 2009. Progress towards DNA barcoding of fungi. Mol Ecol Resour 9:83–89. https://doi.org/10.1111/j.1755-0998.2009.02635.x.
5. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, Karlebach S, Gorle R, Russell J, Tacket CO, Brotman RM, Davis CC, Ault K, Peralta L, Forney LJ. 2011. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A 108:4680–4687. https://doi.org/10.1073/pnas.1002611107.
6. Webster J. 1992. Anamorph-teleomorph relationships, p 99–117. *In* Barlocher F (ed), The ecology of aquatic hyphomycetes. Springer, Berlin, Germany.
7. Seifert KA, Samson RA, Dewaard JR, Houbraken J, Lévesque CA, Moncalvo JM, Louis-Seize G, Hebert PD. 2007. Prospects for fungus identification using CO1 DNA barcodes, with Penicillium as a test case. Proc Natl Acad Sci U S A 104:3901–3906. https://doi.org/10.1073/pnas.0611691104.
8. Rossman A. 2007. Report of the planning workshop for all fungi DNA barcoding. Inoculum 58:1–5.
9. Iwen PC, Hinrichs SH, Rupp ME. 2002. Utilization of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. Med Mycol 40:87–109. https://doi.org/10.1080/mmy.40.1.87.109.
10. Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AF,

Tedersoo L, Vrålstad T, Ursing BM. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytol 166:1063–1068. https://doi.org/10.1111/j.1469-8137.2005.01376.x.
11. White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p 315–322. *In* Innis MA, Gelfand DH, Sninsky JJ, White TJ (ed), PCR protocols: a guide to methods and applications, vol 18. Academic Press, London, United Kingdom.
12. Vilgalys R, Hester M. 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several Cryptococcus species. J Bacteriol 172:4238–4246. https://doi.org/10.1128/jb.172.8.4238-4246.1990.
13. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. 2007. Development of the human infant intestinal microbiota. PLoS Biol 5:e177. https://doi.org/10.1371/journal.pbio.0050177.
14. Glass NL, Donaldson GC. 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. Appl Environ Microbiol 61:1323–1330.
15. Chao A, Colwell RK, Lin CW, Gotelli NJ. 2009. Sufficient sampling for asymptotic minimum species richness estimators. Ecology 90:1125–1133. https://doi.org/10.1890/07-2147.1.
16. Abarenkov K, Henrik Nilsson R, Larsson KH, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AF, Tedersoo L, Ursing BM, Vrålstad T, Liimatainen K, Peintner U, Kõljalg U. 2010. The UNITE database for molecular identification of fungi—recent updates and future perspectives. New Phytol 186:281–285. https://doi.org/10.1111/j.1469-8137.2009.03160.x.
17. 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 Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.

18. Marsland BJ, Gollwitzer ES. 2014. Host-microorganism interactions in lung diseases. Nat Rev Immunol 14:827–835. https://doi.org/10.1038/nri3769.

19. Underhill DM, Iliev ID. 2014. The mycobiota: interactions between commensal fungi and the host immune system. Nat Rev Immunol 14:405–416. https://doi.org/10.1038/nri3684.

20. Cui L, Morris A, Ghedin E. 2013. The human mycobiome in health and disease. Genome Med 5:63. https://doi.org/10.1186/gm467.

21. 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. https://doi.org/10.1111/j.1365-294X.1993.tb00005.x.

22. Martin KJ, Rygiewicz PT. 2005. Fungal-specific PCR primers developed for analysis of the ITS region of environmental DNA extracts. BMC Microbiol 5:28. https://doi.org/10.1186/1471-2180-5-28.

23. Ghannoum MA, Jurevic RJ, Mukherjee PK, Cui F, Sikaroodi M, Naqvi A, Gillevet PM. 2010. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. PLoS Pathog 6:e1000713. https://doi.org/10.1371/journal.ppat.1000713.

24. Taylor DL, McCormick MK. 2008. Internal transcribed spacer primers and sequences for improved characterization of basidiomycetous orchid mycorrhizas. New Phytol 177:1020–1033. https://doi.org/10.1111/j.1469-8137.2007.02320.x.

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

26. Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W, Daza R, Erlich RL, Giannoukos G, Green L, Hollinger A, Hoover CA, Jaffe DB, Juhn F, McCarthy D, Perrin D, Ponchner K, Powers TL, Rizzolo K, Robbins D, Ryan E, Russ C, Sparrow T, Stalker J, Steelman S, Weiand M, Zimmer A, Henn MR, Nusbaum C, Nicol R. 2010. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. Genome Biol 11:R15. https://doi.org/10.1186/gb-2010-11-2-r15.

27. Willger SD, Grim SL, Dolben EL, Shipunova A, Hampton TH, Morrison HG, Filkins LM, O'Toole GA, Moulton LA, Ashare A, Sogin ML, Hogan DA. 2014. Characterization and quantification of the fungal microbiome in serial samples from individuals with cystic fibrosis. Microbiome 2:40. https://doi.org/10.1186/2049-2618-2-40.

28. Op De Beeck MO, 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. https://doi.org/10.1371/journal.pone.0097629.

29. Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspõllu A, Väin E, Saarma I, Salumets A, Donders GG, Metsis M. 2013. Characterization of the vaginal micro- and mycobiome in asymptomatic reproductive-age Estonian women. PLoS One 8:e54379. https://doi.org/10.1371/journal.pone.0054379.

30. Dupuy AK, David MS, Li L, Heider TN, Peterson JD, Montano EA, Dongari-Bagtzoglou A, Diaz PI, Strausbaugh LD. 2014. Redefining the human oral mycobiome with improved practices in amplicon-based taxonomy: discovery of Malassezia as a prominent commensal. PLoS One 9:e90899. https://doi.org/10.1371/journal.pone.0090899.

31. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, Lewis JD, Bushman FD. 2013. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One 8:e66019. https://doi.org/10.1371/journal.pone.0066019.

32. Mukherjee PK, Chandra J, Retuerto M, Sikaroodi M, Brown RE, Jurevic R, Salata RA, Lederman MM, Gillevet PM, Ghannoum MA. 2014. Oral mycobiome analysis of HIV-infected patients: identification of Pichia as an antagonist of opportunistic fungi. PLoS Pathog 10:e1003996. https://doi.org/10.1371/journal.ppat.1003996.

33. Orgiazzi A, Bianciotto V, Bonfante P, Daghino S, Ghignone S, Lazzari A, Lumini E, Mello A, Napoli C, Perotto S, Vizzini A, Bagella S, Murat C, Girlanda M. 2013. 454 pyrosequencing analysis of fungal assemblages from geographically distant, disparate soils reveals spatial patterning and a core mycobiome. Diversity 5:73–98. https://doi.org/10.3390/d5010073.

34. Tang J, Iliev ID, Brown J, Underhill DM, Funari VA. 2015. Mycobiome: approaches to analysis of intestinal fungi. J Immunol Methods 421:112–121. https://doi.org/10.1016/j.jim.2015.04.004.

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

36. Sapkota R, Knorr K, Jørgensen LN, O'Hanlon KA, Nicolaisen M. 2015. Host genotype is an important determinant of the cereal phyllosphere mycobiome. New Phytol 207:1134–1144. https://doi.org/10.1111/nph.13418.

37. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.

38. Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R. 2011. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics 27:1159–1161. https://doi.org/10.1093/bioinformatics/btr087.

39. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303.

40. Thompson K. 1968. Programming techniques: regular expression search algorithm. Commun ACM 11:419–422. https://doi.org/10.1145/363347.363387.

41. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal Barcoding Consortium, Fungal Barcoding Consortium Author List. 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for fungi. Proc Natl Acad Sci U S A 109:6241–6246. https://doi.org/10.1073/pnas.1117018109.

42. Kibbe WA. 2007. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 35:W43–W46. https://doi.org/10.1093/nar/gkm234.

43. Schlecht NF, Burk RD, Nucci-Sack A, Shankar V, Peake K, Lorde-Rollins E, Porter R, Linares LO, Rojas M, Strickler HD, Diaz A. 2012. Cervical, anal and oral HPV in an adolescent inner-city health clinic providing free vaccinations. PLoS One 7:e37419. https://doi.org/10.1371/journal.pone.0037419.

44. Smith BC, Zolnik CP, Usyk M, Chen Z, Kaiser K, Nucci-Sack A, Peake K, Diaz A, Viswanathan S, Strickler HD, Schlecht NF, Burk RD. 2016. Distinct ecological niche of anal, oral, and cervical mucosal microbiomes in adolescent women. Yale J Biol Med 89:277–284.

45. Hercus C. 2009. Novocraft short read alignment package. Novocraft Technologies, Pataling Jaya, Selangor, Malaysia. http://www.novocraft.com.

46. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.

47. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. 2012. PANDAseq: paired-end assembler for Illumina sequences. BMC Bioinformatics 13:31. https://doi.org/10.1186/1471-2105-13-31.

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

49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

50. Team RC. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org.

51. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217.

52. Oksanen J, Kindt R, Legendre P, O'Hara B, Henry M, Stevens H. 2007. The vegan package. Community ecology package 10:631–637. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org.

53. Batdorf CS. January 1903. Coin-package. US patent US717964 A.

54. Wickham H. 2012. reshape2: flexibly reshape data: a reboot of the reshape package. R package version 1. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org.

55. Wickham H, Chang W. 2013. An implementation of the Grammar of Graphics. R package version. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org.